

AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 14, NUMBER 1

JANUARY-FEBRUARY, 1946

Probability, Frequency and Reasonable Expectation

R. T. Cox

The Johns Hopkins University, Baltimore 18, Maryland

Frequency or Reasonable Expectation as the Primary Concept

THE concept of probability has from the beginning of the theory involved two ideas: the idea of frequency in an ensemble and the idea of reasonable expectation. The choice of one or the other as the primary meaning of probability has distinguished the two main schools of thought in the theory.¹

If a box contains two white balls and one black ball, indistinguishable except by color, both schools agree that the probability that a blindfolded man will draw a white ball on a single trial is $\frac{2}{3}$ and the probability that he will draw a black ball is $\frac{1}{3}$. On the frequency theory, the primary meaning of these probabilities is in terms of the ensemble. The ensemble may be an indefinitely large number of such boxes having the same contents, or it may be an indefinitely large number of drawings from the same box, the ball drawn being replaced each time. The significant point is that the initial circumstances are assumed to be capable of indefinite repetition, these repetitions constituting the ensemble. That the probability of a white ball is $\frac{2}{3}$ means simply that the number of

trials giving a white ball as result is $\frac{2}{3}$ the number of trials in the whole ensemble. According to the frequency theory, this is not a prediction of the theory of probability but the definition of the probability. Probability in that theory is a characteristic of the ensemble and, without the ensemble, cannot be said to exist.

Again, according to both schools, the probability of a white ball in two successive drawings, when the first ball drawn is not replaced, is $\frac{2}{3} \times \frac{1}{2}$, or $\frac{1}{3}$. According to the frequency theory, this implies that two balls are drawn successively from each of an ensemble of boxes containing originally two white balls and one black ball. On $\frac{2}{3}$ of the trials a white ball is drawn first and one white and one black ball are left in the box. Then, in $\frac{1}{3}$ of the trials which give this result, a white ball is drawn next, so that $\frac{1}{3}$ of the whole number of trials give white balls on both drawings. These examples illustrate the general fact that, when probability is identified with frequency in an ensemble, the probabilities are calculated by arithmetic in particular examples and, in more general cases, the rules of probability are found by ordinary algebra.²

¹ If minor differences are counted, the number of schools seems to be somewhere between two and the number of authors, and probably nearer the latter number. But the clearest line of division is the one mentioned.

² An exposition of the frequency theory, with some comment on other theories, has been given by G. Bergmann, *Am. J. Physics* 9, 263 (1941). Readers with a wider knowledge of philosophy than mine will be better able to compare his views with those of this paper.

Probability is recognized also as providing a measure of the reasonable expectation of an event in a single trial. That the probability of drawing a white ball is $\frac{2}{3}$ and of drawing a black ball is $\frac{1}{3}$ means that a white ball is a more likely result of a trial than a black ball, and the numbers $\frac{2}{3}$ and $\frac{1}{3}$ serve to compare the likelihoods of the two results. According to the second main school of probability, this measure of reasonable expectation, rather than the frequency in an ensemble, is the primary meaning of probability.

If it could be shown that every measure of reasonable expectation is also a frequency in some ensemble and that every frequency in an ensemble measures a reasonable expectation, then the choice of one or the other as the primary meaning of probability would not be very important. I shall not attempt to discuss whether there are frequencies in an ensemble that are not measures of reasonable expectation. It is enough for my present purpose to show that the two interpretations are not always identical. For this it will suffice to point out that there are probabilities in the sense of reasonable expectations for which no ensemble exists and for which, if one is conceived, it is clearly no more than a convenient mental artifice. Thus, when the probability is calculated that more than one planetary system exists in the universe, it is barely tenable even as an artifice that this refers to the number of universes having more than one planetary system among an indefinitely large number of universes, all resembling in some way the universe, which by definition is all-inclusive.

Moreover, there is so gradual a transition from the cases in which there is a discoverable ensemble and those in which there is none that a theory which requires a sharp distinction between them offers serious difficulties. A few examples will illustrate this point. Let us consider the probability that the number of heads thrown in a certain number of tosses of an unbiased coin shall lie within certain limits, and let us compare with this the probability, often considered, that the true value of a physical constant lies within certain limits. The two probabilities have something in common, but there is a difference between them. The difference lies in the causes that oblige us to deal with probabilities rather than certainties in discussing

the score in tossing a coin and the value of a physical constant. In discussing the score in a given number of tosses of a coin, we have to use probabilities because the score will vary from one trial to another. The true value of a physical constant, on the other hand, is unique. We have to speak of the probability that it lies within certain limits only because our knowledge is incomplete.

Sometimes, it is true, the probability that the value of a physical constant lies within certain limits is equivalent to another probability, that the error in the average of a number of measurements lies within these limits. If there are no systematic sources of error, we may imagine an ensemble of measurements and treat the measurements made as a random sample of this ensemble. The probability in question may then be found in a manner similar to that used in dealing with the coin. For example, the probability of certain limits for the true value of the Joule equivalent may perhaps be considered in this way.

The case is somewhat different with the reciprocal fine-structure constant that appears in quantum mechanics. For here, in addition to the values derived from measurements, there is evidence of another sort in the argument adduced by Eddington³ that this constant may be expected to be an integer, having the value 137. If it should be estimated from the measurements alone that there is an equal probability that the constant lies inside or outside of certain limits which include 137, then Eddington's argument will increase the probability that it lies inside these limits and correspondingly decrease the probability that it lies outside.

As a final example, we may consider the case of a purely mathematical constant, of which the existence has been proved but the value determined only within certain limits. A problem of the theory of numbers, discussed by Hardy⁴ among others, provides a good example. It concerns the equivalence of an integer to a sum

³ A. S. Eddington, *Relativity theory of protons and electrons* (Macmillan, New York, and Univ. Press, Cambridge, 1936).

⁴ G. H. Hardy, *Some famous problems of the theory of numbers and in particular Waring's problem* (Clarendon Press, Oxford, 1920).

of cubes of smaller integers. It has been proved that any integer is given by the sum of not more than 9 cubes, and that any integer beyond a certain one is given by the sum of 8 or fewer. It is expected that large enough integers may all be expressed by the sum of some still smaller number of cubes, and the problem is to find the minimum number required for all integers above a certain value. It has been proved that, if this value is taken high enough, the number in question is 4, 5, 6, 7 or 8. This is as far as rigorous proof has gone, but the evidence of computation makes some of these numbers less likely than others. In very large samples—all of the first 40,000 integers and the 2000 ending at one million—the integers requiring 8 cubes are found to drop out early in the progress to higher integers and those requiring 7 disappear somewhat farther on, while those requiring 6 occur more and more rarely until there are only two among the 2000 integers next below one million. Hardy concludes that the minimum number for large enough integers is almost certainly neither 8 nor 7 and probably not 6. There remain 5 and 4 as the likely numbers, and he seems to favor 4 as the more probable.

Let us consider now these four examples: the probability of certain limits for (i) the score in a number of throws of a coin, (ii) the value of the Joule equivalent, (iii) the value of the reciprocal fine-structure constant and (iv) the value of the least number of cubes for the expression of large integers. It will most likely be granted that other examples can be interpolated among these, so that the differences will be very slight between each example and those next before and after. We shall have then a graded series of examples of probability. At one end of the series the interpretation of probability in terms of frequency will be valid, at the other end it will be impossible. For it is certainly impossible to discuss the statistical spread of the determinations of a number which has never in fact been determined and of which the determination, when it is made, will give a single and logically inevitable value.

Nevertheless, it must be admitted that there is a kind of reasoning common to all these examples. The gambler in the first example, the physicist in the second and third, and the mathe-

matician in the fourth are all using similar processes of inference.

In this connection it is worth while to observe how much of the theory of probability deals with relations between probabilities: between the probability that an event will not occur and the probability that it will occur, between the probability of both of two events and their separate probabilities, between these probabilities and the probability that at least one of the two events will occur. In the case of probabilities that can be identified with frequencies in an ensemble, these relations are readily obtained by ordinary algebra, as was mentioned earlier. But the same or at least similar relations are involved in inference concerned with reasonable expectation even when no ensemble is discoverable. Thus, under any definition of probability, or even without an attempt to define it precisely, there will still be agreement that the less likely an event is to occur the more likely it is not to occur. The occurrence of both of two events will not be more likely and will generally be less likely than the occurrence of the less likely of the two. But the occurrence of at least one of the events is not less likely and is generally more likely than the occurrence of either.

For example, if it were a question of the credibility of a certain hypothesis for the origin of terrestrial life or of human language, one would hold it as a point against the hypothesis that it postulated the occurrence of two events, of which neither was considered very probable, but the hypothesis would gain in credibility if it could be justified by postulating merely that one or the other of these events had occurred. Generally speaking, a simple hypothesis is preferred to a complex one. If this preference is founded on a reasonable belief rather than being a mere convention, its justification would seem to be that two or more postulates are less likely to be true than a single one of about the same likelihood.

This difficulty of the frequency theory of probability may now be summarized. There is a field of probable inference which lies outside the range of that theory. The derivation of the rules of probability by ordinary algebra from the characteristics of the ensemble cannot justify the use of these rules in this outside field.

Nevertheless, the use of these rules in this field is universal and appears to be a fundamental part of our reasoning. Thus the frequency theory is inadequate in the sense that it fails to justify what is conceived to be a legitimate use of its own rules.

From a purely rational point of view, the extent of this field of inference outside the range of the frequency theory is irrelevant to the point in question. Even if the valid instances of reasoning in this field were rare and of little consequence, it would still be logically necessary to maintain the inadequacy of the frequency theory. As a practical matter, however, if these instances were few or trivial, we should probably be content to ignore them. But actually, as I have tried to suggest by the examples given, it is rather the cases in which a strictly definable ensemble exists that are exceptional. This is not to say that they are numerically few. There are many of them, and they have a particular interest, but they still do not appear to comprise in our ordinary practice the greater part of the uses of probable inference. Nor are the other uses by any means trivial. Kemble,⁵ in an interesting paper which covers, among other things, some of the ground thus far traversed here, has made the point that the frequency definition of probability does not suffice to establish the connection between statistical mechanics and thermodynamics, which is certainly crucial in physical theory.

A very original and thoroughgoing development of the theory of probability, which does not depend on the concept of frequency in an ensemble, has been given by Keynes.⁶ In his view, the theory of probability is an extended logic, the logic of probable inference. Probability is a relation between a hypothesis and a conclusion, corresponding to the degree of rational belief and limited by the extreme relations of certainty and impossibility. Classical deductive logic, which deals with these limiting relations only, is a special case in this more general development. Hence it follows in general that the theory of probability cannot be based entirely on concepts of classical logic. In particular,

the relation of probability cannot be defined in terms of certainty, since certainty itself is a special case of probability. The frequency definition of probability is therefore invalid, since it depends on the relations of certainty involved in the knowledge of numbers of instances. Probability is taken as a primary concept, like distance or time in mechanics, not reducible to any more elementary terms.

Merely to describe Keynes' position, as I have done, without giving the reasoning by which he is led to it, does his work very poor justice. The reasoning is, to me at least, very convincing, and is the original source of a large part of the opinions given here, though the arguments I have used are not the same as his. Nevertheless, it must be conceded that his work does not bring us very far in the solution of the problem mentioned earlier, that of justifying the few basic rules of probable inference necessary for the development of the theory. These rules, in Keynes' theory, are simply taken as axiomatic. Now some primary assumptions will have to be made by anyone who accepts, as I am strongly inclined to do, his general point of view as to the nature of probability, because some rational starting point is needed to replace the frequency definition, once that has been abandoned. But Keynes' axioms seem to me, as they have doubtless seemed to others, including Kemble, somewhat too arbitrary and too sophisticated to be entirely suitable as axioms. They do not appeal very directly to common sense, and it is hard to see how they would have been formulated without considering colored balls in a box, dice, coins, or some of the other devices associated with the concept of the ensemble. It is rather as if Euclid had placed the Pythagorean theorem among the axioms of plane geometry.

Russell⁷ makes a criticism somewhat different in form, but which may have the same ground as this. After conceding the strength of Keynes' argument against the frequency theory, he nevertheless prefers that theory, if it can be logically established, because of its explicit definition of probability. It is this definition that makes it possible to avoid the assumption of axioms such as characterize Keynes' theory.

⁵ E. C. Kemble, *Am. J. Physics* 10, 6 (1942).

⁶ J. M. Keynes, *A treatise on probability* (Macmillan, London, 1929).

⁷ B. Russell, *Philosophy* (Norton, New York, 1927).

Other authors, who, like Keynes, present an axiomatic development, choose somewhat different sets of postulates, but those I have seen still show some of the tool marks of their original derivation from the study of games of chance, with the consequent implication of an ensemble. I think this is true even of the carefully chosen postulates of Jeffreys and Wrinch, whether in their original form or as revised by Jeffreys.⁸

Relations of Reasonable Expectation Consistent with Symbolic Logic

In what follows next, I shall try to show that by employing the algebra of symbolic logic it is possible to derive the rules of probability from two quite primitive notions, which are independent of the concept of the ensemble and which, as I think, appeal rather immediately to common sense. This algebra has been applied to probability by a number of writers, including Boole,⁹ who originated it. Still, its possibilities in this respect do not seem to have been fully realized. It may be well here to give a brief introduction to the Boolean algebra, at least to as much of it as the later argument will require.

Letters, a , b , c , \dots , will denote propositions. There is an advantage in speaking of the probabilities of propositions rather than of events, partly for the sake of greater generality but mainly because speaking of events easily invokes the notion of sequence in time, and this may become a source of confusion. A proposition may, of course, assert the occurrence of an event, but it may just as well assert something else, for example, something about a physical constant. The proposition not- a will be denoted by $\sim a$, the proposition a -and- b by $a \cdot b$, and the proposition a -or- b by $a \vee b$.

It is to be borne in mind that the proposition $\sim a$ is not the particular proposition which in some sense is the opposite of a . Thus if a is the proposition, "The stranger was a short, fat old man without coat or hat," $\sim a$ is not the proposition, "The stranger was a tall, thin young woman with coat and hat." To assert $\sim a$ means nothing more than to answer "no" to the ques-

tion, "Is a wholly true?" If a is in several parts, a_1 , a_2 , \dots , to assert $\sim a$ is not to affirm that a_1 , a_2 , \dots are all false but only to say that at least one of them is false.

Since the letters a , b denote propositions and not events, the order in which they appear in the symbols $a \cdot b$ and $a \vee b$ is only the order in which two propositions are stated, not the order in time in which two events occur. Also the form $a \cdot a$ indicates only that a proposition is twice stated, not that an event has twice occurred.

It is also to be understood that $a \vee b$ means a -or- b in the sense of the child who asks, "May I have a nickel or a dime?" without meaning to exclude the possibility of both a nickel and a dime, not in the sense of the orator saying "Sink or swim, survive or perish." Thus $a \vee b$ has the sense for which the form a and/or b is often employed.

Finally it may be noted that if the proposition, "It is raining," is true, then the proposition, "It is raining or snowing," is also true. To assert a proposition a is to imply every proposition $a \vee b$ of which a is one term.

With the meaning of the symbols thus understood, the rules for their combination may be set down as follows:

$$\sim \sim a = a, \quad (1)$$

$$a \cdot b = b \cdot a, \quad (2) \quad a \vee b = b \vee a, \quad (2')$$

$$a \cdot a = a, \quad (3) \quad a \vee a = a, \quad (3')$$

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c = a \cdot b \cdot c, \quad (4)$$

$$a \vee (b \vee c) = (a \vee b) \vee c = a \vee b \vee c, \quad (4')$$

$$\sim (a \cdot b) = \sim a \vee \sim b, \quad (5)$$

$$\sim (a \vee b) = \sim a \cdot \sim b, \quad (5')$$

$$a \cdot (a \vee b) = a, \quad (6) \quad a \vee (a \cdot b) = a. \quad (6')$$

These eleven rules are not all independent. From six of them it is possible to prove the remaining five, and the set of six may be chosen in various ways. It is necessary only to include the first and one from each similarly numbered pair of the others. Thus, for example, Eq. (5') is derived as follows.

$$\begin{aligned} \sim (a \vee b) &= [\text{by Eq. (1)}] \sim (\sim \sim a \vee \sim \sim b) \\ &= [\text{by Eq. (5)}] \sim \sim (\sim a \cdot \sim b) \\ &= [\text{by Eq. (1)}] \sim a \cdot \sim b. \end{aligned}$$

⁸ H. Jeffreys, *Theory of probability* (Clarendon Press, Oxford, 1939).

⁹ G. Boole, *An investigation of the laws of thought* (Macmillan, London, 1854).

Now let the symbol $b|a$ denote some measure of the reasonable credibility of the proposition b when the proposition a is known to be true.¹⁰ The term is indefinite at this point, because, if there is one such measure, there will be any number of others. If $b|a$ is one such measure, then an arbitrary function $f(b|a)$ will also be a measure. Consequently, the symbol is not now to be identified with the conventional probability. To avoid that implication, I shall call $b|a$ the *likelihood* of the proposition b on the hypothesis a , taking advantage of a suggestion made by Margenau,¹¹ but at the same time taking the liberty of giving the term a more inclusive meaning than the one he proposed.

It is not to be supposed that a relation of likelihood exists between any two propositions. If a is the proposition "Caesar invaded Britain" and b is "Tomorrow will be warmer than today," there is no likelihood $b|a$, because there is no reasonable connection between the two propositions.

It is now time to make the first of the two assumptions mentioned earlier as providing a basis for the principles of probable inference. We assume, whatever measure be chosen, that the likelihood $c \cdot b|a$ is determined in some way by the two likelihoods $b|a$ and $c|b \cdot a$, or

$$c \cdot b|a = F(c|b \cdot a, b|a), \quad (7)$$

where F is some function of two variables.

Written in symbolic form, this assumption may not appear very axiomatic. Actually it is a familiar enough rule of common sense, as an example will show. Let b denote the proposition that an athlete can run from one given place to another, and let c denote the proposition that he can run back without stopping. The physical condition of the runner and the topography of the course are described in the hypothesis a . Then $b|a$ is the likelihood that he can run to the distant place, estimated on the information given in a , and $c|b \cdot a$ is the likelihood that he can run back, estimated on the initial information and the further assumption that he has

¹⁰ Keynes has traced the use of such a symbol to H. McColl, *Proc. Lond. Math. Soc.* 11, 113 (1880). McColl uses the symbol x_a for the probability of the proposition x on the hypothesis a .

¹¹ H. Margenau, *Am. J. Physics* 10, 224 (1942). R. A. Fisher has used the term in a quite different sense.

just run one way. These are just the likelihoods that would have to be considered in estimating the likelihood, $c \cdot b|a$, that he can run the complete course without stopping. In postulating only that the last-named likelihood is some function of the other two, we are making the least restrictive assumption possible.

The form of the function F is partly conventional because of the indefiniteness of the measure to be used for likelihood. But it is not wholly so, for it must be consistent with the algebra of propositions. Accordingly we make use of Eq. (4) to derive a functional equation involving F , as follows:

$$\begin{aligned} d \cdot c \cdot b|a &= [\text{by Eq. (4)}] (d \cdot c) \cdot b|a \\ &= [\text{by Eq. (7)}] F(d \cdot c|b \cdot a, b|a). \end{aligned}$$

But

$$\begin{aligned} d \cdot c|b \cdot a &= [\text{by Eq. (7)}] F(d|c \cdot (b \cdot a), c|b \cdot a) \\ &= [\text{by Eq. (4)}] F(d|c \cdot b \cdot a, c|b \cdot a). \end{aligned}$$

Hence

$$d \cdot c \cdot b|a = F[F(d|c \cdot b \cdot a, c|b \cdot a), b|a].$$

Also

$$\begin{aligned} d \cdot c \cdot b|a &= [\text{by Eq. (4)}] d \cdot (c \cdot b)|a \\ &= [\text{by Eq. (7)}] F(d|(c \cdot b) \cdot a, c \cdot b|a) \\ &= [\text{by Eqs. (4) and (7)}] \\ &\quad F[d|c \cdot b \cdot a, F(c|b \cdot a, b|a)]. \end{aligned}$$

Equating these two expressions for $d \cdot c \cdot b|a$ and, for simplicity, letting $d|c \cdot b \cdot a = x$, $c|b \cdot a = y$, and $b|a = z$, we have

$$F[F(x, y, z)] = F[x, F(y, z)]. \quad (8)$$

The function F must be such as to satisfy Eq. (8) for arbitrary values of x , y and z . It is easily shown by substitution that this equation is satisfied if

$$Cf[F(p, q)] = f(p)f(q),$$

where f is an arbitrary function of a single variable, and C is an arbitrary constant. It is shown in the appendix that this is also the general solution, provided F has continuous second derivatives. We have then

$$Cf(c \cdot b|a) = f(c|b \cdot a)f(b|a).$$

The choice of the function f is purely a matter of convention. For it has already been pointed out that, if $b|a$ is a measure of the credibility of

b on the hypothesis a , then so also is $f(b|a)$. We might then continue the discussion with $f(b|a)$ as the symbol of likelihood in place of $b|a$ and never have to specify the function f . But this would give two symbols where one would be enough. As a matter of convenience, therefore, we write

$$Cc \cdot b|a = c|b \cdot a \cdot b|a. \quad (9)$$

This is, of course, the same as choosing the function f to make $f(b|a) = b|a$. Since the choice was conventional, it follows that another choice could have been made. We might, for example, have let $f(b|a) = \exp(b|a)$, whence it would have followed that the likelihood of $c \cdot b$ was, except for an arbitrary additive constant, equal to the sum of the likelihoods which determine it. This would have given us a likelihood related to the one we have as entropy is related to thermodynamic probability in statistical mechanics. It would have been an allowable choice, but a less convenient one than that which was made.

If in Eq. (9) we let $c = b$, and note that $b \cdot b = b$, by Eq. (3), we obtain, after dividing by $b|a$,

$$C = b|b \cdot a.$$

Thus we see that when the hypothesis includes the conclusion the likelihood has the constant value C , whatever the propositions may be. This is what we should expect, since b is certain on the hypothesis $b \cdot a$, and we do not recognize degrees of certainty.

The value to be assigned to C , the likelihood of certainty, is purely conventional. If it is desired to make the likelihoods with which we are dealing correspond as nearly as possible to ordinary probabilities, then C will be given the value 1. Other choices are often made, especially in conversation. The phrase "one chance in a hundred" may be taken to mean unit likelihood on a scale in which certainty is represented by 100. Statements that have the form of assertions about numbers in an ensemble may be merely convenient ways of stating likelihoods on a scale chosen for its aptness to the question considered. In a general discussion the most convenient value for C is unity, and we therefore write Eq. (9) in the form

$$c \cdot b|a = c|b \cdot a \cdot b|a. \quad (10)$$

This has the same form as the ordinary rule for the probability of two events. However, it does not make our likelihood correspond uniquely to the ordinary probability. For Eq. (10) raised to any power m is

$$(c \cdot b|a)^m = (c|b \cdot a)^m (b|a)^m.$$

Thus any power of our likelihood satisfies an equation of the same form as Eq. (10) and corresponds equally well to the ordinary probability.

Next to be sought is a second assumption of probable inference, which is to provide a relation between the likelihoods of the propositions b and $\sim b$ on the same hypothesis a . Since $\sim b$ is determined when b is specified, a reasonable assumption, and the least restrictive possible, appears to be that $\sim b|a$ is determined by $b|a$, or

$$\sim b|a = S(b|a), \quad (11)$$

where S is some function of a single variable.

By Eq. (1), $\sim \sim b|a = b|a$, and therefore $S[S(b|a)] = b|a$. Thus S must be such a function that

$$S[S(x)] = x, \quad (12)$$

where x may have any possible value of a likelihood between those of certainty and impossibility. This does not impose enough restriction on S to be of much use by itself. Another functional equation may be obtained by considering $S(c \vee b|a)$; thus,

$$S(c \vee b|a) = \sim(c \vee b)|a \\ = [\text{by Eq. (5')}] \sim c \cdot \sim b|a.$$

We wish to eliminate the propositions $\sim c$ and $\sim b$, so as to obtain an equation in the propositions c , b and a and the function S . First we eliminate $\sim c$.

$$\sim c \cdot \sim b|a = [\text{by Eq. (10)}] \sim c| \sim b \cdot a \sim b|a \\ = [\text{by Eq. (11)}] S(c| \sim b \cdot a) S(b|a).$$

Thus we have

$$S(c \vee b|a) = S(c| \sim b \cdot a) S(b|a),$$

or

$$S(c| \sim b \cdot a) = S(c \vee b|a) / S(b|a).$$

Taking the function S of both sides of this equation and using Eq. (12), we obtain

$$c| \sim b \cdot a = S[S(c \vee b|a) / S(b|a)]. \quad (13)$$

Next we eliminate $\sim b$:

$$\begin{aligned} c|\sim b \cdot a &= [\text{by Eq. (10)}] c \cdot \sim b|a / \sim b|a \\ &= [\text{by Eq. (2)}] \sim b \cdot c|a / \sim b|a \\ &= [\text{by Eq. (10)}] \sim b \cdot c \cdot a \cdot c|a / \sim b|a \\ &= [\text{by Eq. (11)}] S(b|c \cdot a)c|a / S(b|a). \end{aligned}$$

Therefore we may write in place of Eq. (13),

$$S(b|c \cdot a)c|a / S(b|a) = S[S(c \vee b|a) / S(b|a)].$$

It is convenient now to have a as the common hypothesis in all the likelihoods. We note that

$$\begin{aligned} b|c \cdot a &= [\text{by Eq. (10)}] b \cdot c|a / c|a \\ &= [\text{by Eq. (2)}] c \cdot b|a / c|a. \end{aligned}$$

Substituting this expression in the preceding equation and multiplying both sides by $S(b|a)$, we obtain

$$\begin{aligned} S(c \cdot b|a / c|a)c|a \\ = S[S(c \vee b|a) / S(b|a)]S(b|a). \quad (14) \end{aligned}$$

This equation must hold for arbitrary meanings of the propositions a , b and c . Let $b=c \cdot d$. Then

$$c \vee b = c \vee (c \cdot d) = [\text{by Eq. (6')}] c,$$

and

$$\begin{aligned} c \cdot b &= c \cdot (c \cdot d) \\ &= [\text{by Eq. (4)}] (c \cdot c) \cdot d = [\text{by Eq. (3)}] c \cdot d. \end{aligned}$$

Making these substitutions in Eq. (14), we obtain

$$S(c \cdot d|a / c|a)c|a = S[S(c|a) / S(c \cdot d|a)]S(c \cdot d|a).$$

This may be written in a highly symmetric form if we let $c|a=x$ and $S(c \cdot d|a)=y$, and make use of the fact that $c \cdot d|a=[\text{by Eq. (12)}] S[S(c \cdot d|a)]=S(y)$. In these terms we have

$$xS[S(y)/x]=yS[S(x)/y]. \quad (15)$$

This equation must be satisfied by the function S for all of the values of x and y obtainable by arbitrarily varying the propositions c , d and a . If the function S is twice differentiable, the solution of Eq. (15) together with Eq. (12) is, as shown in the appendix,

$$S(p)=(1-p^m)^{1/m},$$

where m is an arbitrary constant. Hence by Eq. (11),

$$(b|a)^m + (\sim b|a)^m = 1.$$

Now, whatever the value of m , if $b|a$ measures the credibility of b on the hypothesis a , then so

also will $(b|a)^m$. It has already been pointed out that $(b|a)^m$ may replace $b|a$ in Eq. (10). Therefore we may take $(b|a)^m$ as the symbol of likelihood without being under any necessity of assigning a value to m . This is the same as to say that the choice of a value for m is purely conventional. For simplicity of notation we let $m=1$ and write

$$b|a + \sim b|a = 1. \quad (16)$$

This has the same form as the ordinary rule relating the probability of $\sim b$ to that of b , or, as it is usually said, the rule for the probability that an event will not occur, given the probability that it will occur.

If in Eq. (16) we let $b=a$, then

$$a|a + \sim a|a = 1.$$

The two likelihoods are now those of certainty and impossibility. Since certainty has been given the likelihood 1, it now follows that impossibility has the likelihood zero.

Two other useful theorems are easily obtained. By Eq. (10),

$$c \cdot b|a + \sim c \cdot b|a = (c|b \cdot a + \sim c|b \cdot a)b|a.$$

By Eq. (16),

$$c|b \cdot a + \sim c|b \cdot a = 1.$$

Therefore,

$$c \cdot b|a + \sim c \cdot b|a = b|a, \quad (17)$$

This is one of the theorems. The other is obtained as follows.

$$\begin{aligned} c \vee b|a &= [\text{by Eq. (16)}] 1 - \sim (c \vee b)|a \\ &= [\text{by Eq. (5')}] 1 - \sim c \cdot \sim b|a \\ &= [\text{by Eq. (17)}] 1 - \sim b|a + c \cdot \sim b|a. \end{aligned}$$

Now, by Eq. (16), $1 - \sim b|a = b|a$. Also,

$$\begin{aligned} c \cdot \sim b|a &= [\text{by Eq. (2)}] \sim b \cdot c|a \\ &= [\text{by Eq. (17)}] c|a - b \cdot c|a \\ &= [\text{by Eq. (2)}] c|a - c \cdot b|a. \end{aligned}$$

Therefore,

$$c \vee b|a = c|a + b|a - c \cdot b|a, \quad (18)$$

which has the same form as the ordinary rule for the probability that at least one of two events will occur. When, as is often done, the rule is stated for mutually exclusive events, the last term in the right-hand member does not appear. This conceals the rather interesting symmetry of

the equation between the propositions $c \vee b$ and $c \cdot b$.

The indefiniteness of the concept of likelihood, defined only as a measure of reasonable credibility, has been removed by the conventions which have been adopted. The symbol $b|a$ stands now for a particular measure of credibility. Since this measure has been shown to be subject to the ordinary rules of probability, it is appropriate to call it the probability of the proposition b on the hypothesis a , discarding the term likelihood, which was less definitely defined.

The rules obtained, being only relations between probabilities, do not of themselves assign numerical values to all the probabilities arising in specific problems. The only numerical values thus far obtained are those corresponding to certainty and impossibility, and these were assigned by convention rather than required by the rules of symbolic logic. It is hardly to be supposed that every reasonable expectation should have a precise numerical value. In a number of cases, however, the familiar rule of insufficient reason may be employed. If there are n propositions of which, with respect to a given hypothesis, one and no more than one can be true, and if the hypothesis gives no reason for considering any one of them more likely than another, then, by the rules obtained, each of them has the probability $1/n$.

Probability and Frequency

The whole discussion thus far has consisted of two parts. The first part was intended to show that the rules of probable inference are credited by common sense with a wider validity than can be established by deducing them from the frequency definition of probability. In the second part they were derived without reference to this definition, from rather elementary postulates. It remains now to see what is the connection between probability, as here understood, and the frequency of an event.

Let us suppose that two capsules contain equal masses of radon, but that the contents are of different ages, one having been produced by the very recent decay of radium, and the other having been drawn from a vessel in which radon has been accumulating for a long time over radium in solution. Suppose there are two

identical ion counters, each receiving radiation from one capsule, and each placed with respect to its capsule in the same relative position as the other. One of the two capsules will be the first to cause 1000 discharges in its ion counter. More than one hypothesis will ascribe to each capsule the same probability of being first. A physicist will estimate equal probabilities on the ground of the many observations which have been made on rates of radioactive transformation, with some additional evidence from quantum mechanics that the stability of such an aggregate of elementary particles as an atomic nucleus is independent of its age. Another person, quite unfamiliar with all this, will estimate equal probabilities merely on the ground that he does not know which capsule contains the older radon and has therefore no possible reason to suppose that one sample rather than the other will be the first to cause 1000 discharges.

These are the extreme cases, the first estimate being highly significant and the second quite trivial, but each is right on the hypothesis given. Kemble calls the estimate of the physicist one of *objective* probability and that of the other person one of *subjective*, or *primary*, probability. The latter term seems preferable to me, as it does to him. It is true that the estimate of the nonphysicist is subjective in the sense that it is relative to his limited information, but it is objective in the sense that another person with the same information would reasonably make the same estimate. It seems questionable whether there is a real difference in the *kind* of judgment made by the nonphysicist and the physicist. The nonphysicist bases his estimate on the fact that the capsules are indistinguishable. The physicist bases *his* estimate on the accumulated evidence that the atoms of radon themselves are indistinguishable. The difference seems to be not so much a difference in the nature of the evidence as in its amount and relevance or, to use Keynes' suggestive term, its weight.

Now let the experiment with the radon capsules be tried a number of times, with the old and new samples identified in advance of each trial. Even a long run of instances in which the older sample is first will not change the probabilities as estimated by the physicist. The evidence on which he made his first estimate had so much

weight that no additional number of instances, not enormously large, could require a new estimate. The probabilities have for practical purposes become stable. Strictly speaking, since probability is relative to an experience that is never complete, it is always subject to change by new experience. A stable probability is a limit that is not strictly attainable, but that can in certain cases be approximated as nearly as necessary for practical use. It is to be expected that a stable probability will give a better basis for prediction than will an unstable one.

Let \mathbf{a} be a hypothesis of which a number of instances may be examined. Let \mathbf{b}_r mean that a certain proposition \mathbf{b} is valid in the r th instance of \mathbf{a} . Unless the hypothesis \mathbf{a} itself assigns a stable probability to \mathbf{b} , then $\mathbf{b}_s|\mathbf{a}, \mathbf{b}_s|\mathbf{a} \cdot \mathbf{b}_r$, and $\mathbf{b}_s|\mathbf{a} \cdot \sim \mathbf{b}_r$, will generally all be different; the knowledge that \mathbf{b} is valid or that it is not valid in one instance will affect the reasonable expectation of its validity in another instance. But now let there be included in the hypothesis \mathbf{a} a proposition \mathbf{p} , which asserts that the probability is stable and equal to p , some number between 0 and 1. This means that the probability of \mathbf{b}_s is the same whether \mathbf{b}_s or $\sim \mathbf{b}_s$, or neither is included in the hypothesis. Thus

$$\mathbf{b}_s|\mathbf{a} \cdot \mathbf{p} \cdot \mathbf{b}_r = \mathbf{b}_s|\mathbf{a} \cdot \mathbf{p} \cdot \sim \mathbf{b}_r = \mathbf{b}_s|\mathbf{a} \cdot \mathbf{p} = p.$$

Then by Eqs. (10) and (16) we obtain

$$\begin{aligned}\mathbf{b}_s \cdot \mathbf{b}_r |\mathbf{a} \cdot \mathbf{p} &= p^2, \\ \mathbf{b}_s \cdot \sim \mathbf{b}_r |\mathbf{a} \cdot \mathbf{p} &= p(1-p), \\ \sim \mathbf{b}_s \cdot \sim \mathbf{b}_r |\mathbf{a} \cdot \mathbf{p} &= (1-p)^2.\end{aligned}$$

Let n_N mean that the number of instances of \mathbf{b} in N instances of \mathbf{a} is exactly n . Then by Eqs. (10), (16) and (18) it is possible to derive the well-known result of Bernoulli, that

$$n_N |\mathbf{a} \cdot \mathbf{p} = p^n(1-p)^{N-n}N!/n!(N-n)!.$$

This is a maximum when $p=n/N$, and the maximum becomes sharper as N is increased. Thus, when there is a stable probability, the frequency may confidently be expected to approach it as a limit.

There will sometimes be questions in which the existence of a stable probability is known but its value is undetermined. As a rather artificial but simple example, let it be supposed that there are two dice, both dynamically symmetric, but one

of them defectively marked, having two faces instead of one stamped with four dots. Then for either of these dice there is a stable probability of throwing a four, equal to $\frac{1}{6}$ if it is the true die and to $\frac{1}{3}$ if it is the defective one. Suppose one die of the pair is picked up at random and, without being examined, is tossed N times. If a four turns up on n of these throws, what is the probability of a four on the next throw?

The problem may be generalized as follows. Let it be supposed that in the ensemble of instances of a proposition \mathbf{a} , another proposition \mathbf{b} is known to have a stable probability, but the value of this stable probability is unknown. As before, \mathbf{p} will denote the proposition that the probability is stable and equal to a number p , but in the present case \mathbf{p} is not a part of the hypothesis. Instead, the hypothesis contains a weaker proposition which only assigns a probability to the proposition \mathbf{p} corresponding to every value of p .

We may let the single symbol \mathbf{a} represent the entire initial hypothesis, including this proposition. Thus, in the example of the dice, \mathbf{a} will describe the two dice and will also assert that one is chosen at random and tossed without being identified. There are then only two possible stable probabilities, $\frac{1}{6}$ and $\frac{1}{3}$, and they are equally probable at the beginning. Hence in this example $\mathbf{p}|\mathbf{a}$ has the value $\frac{1}{2}$ if p is either $\frac{1}{6}$ or $\frac{1}{3}$ and the value zero if p is any other number.

Returning to the general problem, we suppose that N instances of \mathbf{a} are observed, and \mathbf{b} is found valid in n of them and invalid in the rest. What is now the probability of \mathbf{b} in the $N+1$ th instance of \mathbf{a} ? As before, let n_N denote the proposition that n is the number of instances of \mathbf{b} in N instances of \mathbf{a} . The problem is to find $\mathbf{b}_{N+1}|\mathbf{a} \cdot n_N$, given n and N , and also $\mathbf{p}|\mathbf{a}$ for every value of p .

The theorems available are enough to give the result

$$\mathbf{b}_{N+1}|\mathbf{a} \cdot n_N = \frac{\sum p^{n+1}(1-p)^{N-n}\mathbf{p}|\mathbf{a}}{\sum p^n(1-p)^{N-n}\mathbf{p}|\mathbf{a}},$$

where the summations are over-all values of p .

If, on the hypothesis \mathbf{a} , the stable probability has a continuous range of possible values from 0 to 1, and if $f(p)dp$ denotes the probability of a value between p and $p+dp$, the summations are

replaced by integrals, and we have

$$\mathbf{b}_{N+1}|\mathbf{a} \cdot \mathbf{n}_N = \frac{\int_0^1 p^{n+1}(1-p)^{N-n}f(p)dp}{\int_0^1 p^n(1-p)^{N-n}f(p)dp}$$

It was assumed by Laplace that an unknown probability is equally likely to have any value from 0 to 1. On this assumption, $f(p)$ is constant in the last equation. The integrals in this case are known; and the result, sometimes called the rule of succession, is simply

$$\mathbf{b}_{N+1}|\mathbf{a} \cdot \mathbf{n}_N = (n+1)/(N+2),$$

or approximately, for large values of n and N ,

$$\mathbf{b}_{N+1}|\mathbf{a} \cdot \mathbf{n}_N = n/N.$$

Several authors have pointed out that an unknown stable probability is not necessarily one for which all values from 0 to 1 are equally likely, and the rule of succession has been shown to lead to some absurd results. Nevertheless, we should expect that for large numbers it will generally be right. For we know, by the theorem of Bernoulli given earlier, that, when there is a stable probability, the ratio n/N is very likely to be nearly equal to it when N is large. Also we have understood a stable probability to be the limit that the probability approaches as the weight of the evidence is increased, and usually the surest way to increase the weight of evidence is to increase the number of observed instances. If the ratio n/N and the probability approach a common limit, then certainly they must approach each other.

If the absurdities to which Laplace's rule has led are examined, they are found to fall into three classes: those in which N is not a very large number, those in which $n/N=1$, and those in which $n/N=0$. (The last two are really one class, since to say that \mathbf{b} is valid in N out of N instances of \mathbf{a} is the same as to say that $\sim\mathbf{b}$ is valid in none of N instances.) If these conditions are excluded, Laplace's rule may be derived from a much less drastic assumption than the assumption that all values of the stable probability are equally likely.

If from the general equation for $\mathbf{b}_{N+1}|\mathbf{a} \cdot \mathbf{n}_N$ we

eliminate n by letting $n/N=\nu$, we obtain

$$\mathbf{b}_{N+1}|\mathbf{a} \cdot \mathbf{n}_N = \frac{\int_0^1 p[p^\nu(1-p)^{1-\nu}]^N f(p)dp}{\int_0^1 [p^\nu(1-p)^{1-\nu}]^N f(p)dp}.$$

If $0 < \nu < 1$, then $p^\nu(1-p)^{1-\nu}$ has a maximum value when $p=\nu$. The N th power of this expression, when N is large enough, will have so pronounced a maximum that its values when p is more than slightly different from ν will be relatively negligible. Hence, unless $f(p)$ is extremely small when $p=\nu$, its only values of importance in the integrals will be those for which p and ν are very nearly equal. Therefore, unless $f(p)$ is rapidly varying around this point, it may be replaced in the integrals by the constant $f(\nu)$.

Thus we arrive again at Laplace's rule. Its generality is much less than Laplace supposed. But it serves to show how a probability approaches stability as the number of instances is increased, and this is all we should expect of it.¹²

* * *

Professor K. O. Friedrichs, of New York University, read a preliminary draft of this paper. I wish to thank him for this kindness and for his help in correcting some mathematical inaccuracies. He is not responsible, of course, for any errors that remain or for the opinion expressed as to the nature of probability.

Appendix: The Solution of the Functional Equations

The first equation to be solved is

$$F[F(x, y), z] = F[x, F(y, z)]. \quad (8)$$

Let $F(x, y) = u$, and let $F(y, z) = v$. Then Eq. (8) becomes $F(u, z) = F(x, v)$. Differentiating this with respect to x , y and z in turn, and writing $F_1(p, q)$ for $\partial F(p, q)/\partial p$ and $F_2(p, q)$ for $\partial F(p, q)/\partial q$, we obtain

$$F_1(u, z)\partial u/\partial x = F_1(x, v), \quad (19)$$

$$F_1(u, z)\partial u/\partial y = F_2(x, v)\partial v/\partial y, \quad (20)$$

$$F_2(u, z) = F_2(x, v)\partial v/\partial z. \quad (21)$$

Differentiating Eq. (20) with respect to x , y and z in turn,

¹² The problem of inverse probability when $n/N=1$ (or 0), which is important in the application of probability to inductive reasoning, is discussed at length by Jeffreys in reference 8.

writing $F_{11}(p, q)$ for $\partial F_1(p, q)/\partial p$, and similarly representing the other second derivatives, we obtain

$$F_{11}(u, z)(\partial u/\partial x)(\partial v/\partial y) + F_1(u, z)\partial^2 u/\partial x \partial y = F_{12}(x, v)\partial v/\partial y, \quad (22)$$

$$F_{11}(u, z)(\partial u/\partial y)^2 + F_1(u, z)\partial^2 u/\partial y^2 = F_{22}(x, v)(\partial v/\partial y)^2 + F_2(x, v)\partial^2 v/\partial y^2, \quad (23)$$

$$F_{12}(u, z)\partial u/\partial y = F_{22}(x, v)(\partial u/\partial y)(\partial v/\partial z) + F_2(x, v)\partial^2 v/\partial y \partial z. \quad (24)$$

Differentiating Eq. (19) with respect to z , or Eq. (21) with respect to x , we obtain

$$F_{12}(u, z)\partial u/\partial x = F_{12}(x, v)\partial v/\partial z. \quad (25)$$

Among Eqs. (20), (22), ..., (25) we can now eliminate the functions of u and v other than their derivatives. Thus, eliminating $F_{12}(u, z)$ and $F_{12}(x, v)$ among Eqs. (22), (24) and (25), we find

$$\begin{aligned} & [F_{11}(u, z)(\partial u/\partial y)^2 - F_{22}(x, v)(\partial v/\partial y)^2](\partial u/\partial x)(\partial v/\partial z) \\ & = F_2(x, v)(\partial^2 v/\partial y \partial z)(\partial v/\partial y)(\partial u/\partial x) \\ & \quad - F_1(u, z)(\partial^2 u/\partial x \partial y)(\partial u/\partial y)(\partial v/\partial z) \end{aligned}$$

Combining this with Eq. (23), we can eliminate $F_{11}(u, z)$ and $F_{22}(x, v)$ together, obtaining

$$F_1(u, z)(\partial v/\partial z)[(\partial^2 u/\partial y^2)(\partial u/\partial x) - (\partial^2 u/\partial x \partial y)(\partial u/\partial y)] = F_2(x, v)(\partial u/\partial x)[(\partial^2 v/\partial y^2)(\partial v/\partial z) - (\partial^2 v/\partial y \partial z)(\partial v/\partial y)].$$

Combining this with Eq. (20), we eliminate $F_1(u, z)$ and $F_2(x, v)$ together and obtain

$$\frac{\partial^2 u/\partial x \partial y}{\partial u/\partial x} - \frac{\partial^2 u/\partial y^2}{\partial u/\partial y} = \frac{\partial^2 v/\partial y \partial z}{\partial v/\partial z} - \frac{\partial^2 v/\partial y^2}{\partial v/\partial y}.$$

This may be written in the form,

$$\frac{\partial}{\partial y} \ln \left(\frac{\partial u/\partial x}{\partial u/\partial y} \right) = - \frac{\partial}{\partial y} \ln \left(\frac{\partial v/\partial z}{\partial v/\partial z} \right).$$

Now $u = F(x, y)$ and $v = F(y, z)$, so that

$$\frac{\partial u/\partial x}{\partial u/\partial y} = \frac{F_1(x, y)}{F_2(x, y)}, \quad \text{and} \quad \frac{\partial v/\partial z}{\partial v/\partial y} = \frac{F_1(y, z)}{F_2(y, z)}.$$

We have then,

$$\frac{\partial}{\partial y} \ln \left[\frac{F_1(x, y)}{F_2(x, y)} \right] = - \frac{\partial}{\partial y} \ln \left[\frac{F_1(y, z)}{F_2(y, z)} \right].$$

Since x appears only in the left-hand member and z only in the right-hand member of this equation, it follows that each member is a function only of the remaining variable y . It will be convenient to denote the integral of this function by $\ln \Phi(y)$, so that we have

$$\frac{\partial}{\partial y} \ln \left[\frac{F_1(x, y)}{F_2(x, y)} \right] = \frac{d}{dy} \ln \Phi(y), \quad (26)$$

and

$$\frac{\partial}{\partial y} \ln \left[\frac{F_1(y, z)}{F_2(y, z)} \right] = - \frac{d}{dy} \ln \Phi(y). \quad (27)$$

Permuting x , y and z in Eq. (27), we obtain

$$\frac{\partial}{\partial x} \ln \left[\frac{F_1(x, y)}{F_2(x, y)} \right] = - \frac{d}{dx} \ln \Phi(x). \quad (28)$$

Multiplying Eq. (28) by dx and Eq. (26) by dy and adding,

we obtain

$$\begin{aligned} \frac{\partial}{\partial x} \ln \left[\frac{F_1(x, y)}{F_2(x, y)} \right] dx + \frac{\partial}{\partial y} \ln \left[\frac{F_1(x, y)}{F_2(x, y)} \right] dy \\ = - d \ln \Phi(x) + d \ln \Phi(y). \end{aligned}$$

The left-hand member being now a complete differential, we may integrate and so find

$$F_1(x, y)/F_2(x, y) = h \Phi(y)/\Phi(x), \quad (29)$$

where h is a constant of integration.

To make use of this result, we divide Eq. (20) by Eq. (21), obtaining

$$\frac{F_1(u, z)}{F_2(u, z)} \frac{\partial u}{\partial y} = \frac{\partial v}{\partial z}.$$

The right-hand member is simply $F_1(y, z)/F_2(y, z)$, and, with the aid of Eq. (29), the equation may be written as

$$\frac{\Phi(z)}{\Phi(u)} \frac{\partial u}{\partial y} = \frac{\Phi(z)}{\Phi(y)}.$$

Replacing in this equation u by its value $F(x, y)$ we have

$$dF(x, y)/\partial y = \Phi[F(x, y)]/\Phi(y). \quad (30)$$

Similarly, from Eqs. (19) and (20) we obtain

$$dF(y, z)/\partial y = \Phi[F(y, z)]/\Phi(y),$$

which becomes, when x and y are written for y and z ,

$$dF(x, y)/\partial x = \Phi[F(x, y)]/\Phi(x). \quad (31)$$

Combining Eqs. (30) and (31) to obtain the differential dF (the variables being understood as x and y) we find

$$dF/\Phi(F) = dx/\Phi(x) + dy/\Phi(y).$$

If we denote $\int [d\phi/\Phi(\phi)]$ by $\ln f(\phi)$, we obtain, by integrating and taking the exponentials of both members of this equation,

$$Cf(F) = f(x)f(y),$$

where C is a constant of integration. This then is the solution of Eq. (8).

The solution of the equation

$$xS[S(y)/x] = yS[S(x)/y] \quad (15)$$

is obtained in a similar manner, but more quickly. This equation and the three derived from it by differentiation with respect to x , to y , and to x and y may be written as follows, when $S(y)/x$ is denoted by u and $S(x)/y$ by v :

$$xS(u) = yS(v), \quad (32)$$

$$uS'(u) - S(u) = -S'(v)S'(x), \quad (33)$$

$$S'(u)S'(y) = -vS'(v) + S(v), \quad (34)$$

$$uS''(u)S'(y)/x = vS''(v)S'(x)/y. \quad (35)$$

Multiplying Eq. (32) by Eq. (35), we eliminate x and y simultaneously, obtaining

$$uS''(u)S(u)S'(y) = vS''(v)S(v)S'(x).$$

With this equation, together with Eqs. (33) and (34), it is possible to eliminate $S'(x)$ and $S'(y)$. The result is the equation

$$\frac{uS''(u)S(u)}{[uS'(u) - S(u)]S'(u)} = \frac{vS''(v)S(v)}{[vS'(v) - S(v)]S'(v)}.$$

Since each member of the foregoing equation is the same function of a different variable, this function must

be equal to a constant. Calling this constant k , we have

$$uS''(u)S(u) = k[uS'(u) - S(u)]S'(u).$$

This may be put in the form

$$dS'/S' = k(dS/S - du/u),$$

whence, by integration,

$$S' = A(S/u)^k,$$

where A is a constant.

The variables being separable, another integration gives

$$S^m = Au^m + B,$$

where m has been written for $1-k$, and B is a constant of integration. It is now found by substitution that Eq. (15) can be satisfied for arbitrary values of x and y only if $B = A^2$. Finally, if the solution of Eq. (15) is also to satisfy the equation $S[S(x)] = x$, it is necessary that $A = -1$. Thus we obtain $[S(u)]^m + u^m = 1$.

Application of Group Theory to the Calculation of Vibrational Frequencies of Polyatomic Molecules¹

ARNOLD G. MEISTER AND FORREST F. CLEVELAND
Illinois Institute of Technology, Chicago, Illinois

WILSON² has devised a method for obtaining the vibrational frequencies of polyatomic molecules in which group theory is used to simplify the calculations. The method is especially good for molecules having considerable symmetry and several equivalent atoms, that is, atoms with identical nuclei that transform into one another for all operations of the point group of the molecule. A further advantage of the method is that it requires no coordinate system, but only bond distances, interbond angles and unit vectors directed along the bonds.

Since one who is beginning calculations of vibrational frequencies may find the symbolism of Wilson's papers difficult, and since other papers involving the method omit many of the details, it seems worth while to give an elementary treatment of a few typical molecules for those desiring to start work in this field. The H₂O molecule is considered first because it has only a small number of atoms, has no degenerate frequencies, and permits the reader to concentrate on the method without being confused by the complexity of the molecule.³ Then the CH₃Cl, CH₄ and CD₄ molecules are treated to show how the method is applied when doubly or triply degenerate frequencies are present.⁴

THE H₂O MOLECULE

Symmetry Coordinates

The methods given in a previous paper⁵ are used to determine the point group of the molecule as well as the number of fundamental vibrations of each type. It is found that the H₂O molecule belongs to the point group C_{2v} and that there are two vibrations of type A₁ and one of type B₂. Since a nonlinear molecule containing N atoms has $3N-6$ vibrational degrees of freedom, $3N-6$ coordinates are necessary to describe the vibrations of the molecule. To attain the simplification made possible by the use of group theory, it is

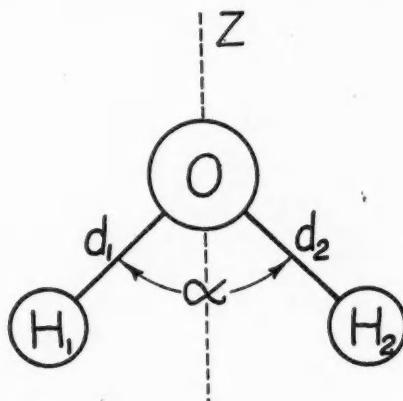


FIG. 1. Bond distances d_1 and d_2 , interbond angle α , and principal symmetry axis Z for the H₂O molecule.

¹ Communication No. 43 from the Spectroscopy Laboratory.

² E. B. Wilson, Jr., *J. Chem. Physics* 7, 1047 (1939); 9, 76 (1941).

³ A more complicated molecule, CH₃Cl, involving only nondegenerate frequencies has been discussed by G. Glockler, *Rev. Mod. Physics* 15, 125 (1943).

⁴ A treatment in outline form of the CH₃Cl molecule is given at the end of Wilson's second paper, reference 2.

⁵ A. G. Meister, F. F. Cleveland and M. J. Murray, *Am. J. Physics* 11, 239 (1943).

TABLE I. Characters for the point group C_{2v} .

C_{2v}	E	C_2	σ_v	σ_v'
A_1	1	1	1	1
B_2	1	-1	-1	1
A_2	1	1	-1	-1
B_1	1	-1	1	-1

necessary that these $3N-6$ coordinates be *symmetry coordinates* which are linear combinations of the *internal coordinates*. The internal coordinates are the changes in bond distances and the changes in interbond angles. It is convenient to construct the symmetry coordinates from equivalent internal coordinates only. Moreover, the choice of linear combinations is not arbitrary but must be made in such a way that the symmetry coordinate transforms according to the characters for the vibration type concerned. Also, the symmetry coordinates must be normalized and orthogonal.

For example, for the H_2O molecule, $3(3)-6$, or three, symmetry coordinates are required. The internal coordinates (see Fig. 1) are Δd_1 and Δd_2 , the changes in the bond distances d_1 and d_2 and $\Delta\alpha$, the change in the interbond angle α . In this case the equivalent internal coordinates are Δd_1 and Δd_2 . Since there are two A_1 vibrations and one B_2 vibration, a suitable set of symmetry coordinates would be: for the A_1 vibrations,

$$R_1 = (1/\sqrt{2})\Delta d_1 + (1/\sqrt{2})\Delta d_2, \quad (1)$$

$$R_2 = \Delta\alpha; \quad (2)$$

and for the B_2 vibration,

$$R_3 = (1/\sqrt{2})\Delta d_1 - (1/\sqrt{2})\Delta d_2. \quad (3)$$

These are of the form

$$R_j = \sum_k U_{jk} r_k, \quad (4)$$

where R_j is the j th symmetry coordinate ($j=1, 2, \dots, 3N-6$), U_{jk} is the coefficient of the k th internal coordinate r_k , and the summation is taken over all of the equivalent internal coordinates. [Eq. (4) can be written more concisely in the matrix notation as $\mathbf{R} = \mathbf{U}\mathbf{r}$.]

The condition for normalization of the j th symmetry coordinate is that

$$\sum_k (U_{jk})^2 = 1. \quad (5)$$

Thus for R_1 , R_2 and R_3 , one has

$$\text{and } (1/\sqrt{2})^2 + (1/\sqrt{2})^2 = 1, \quad (1)^2 = 1,$$

$$(1/\sqrt{2})^2 + (-1/\sqrt{2})^2 = 1;$$

so R_1 , R_2 and R_3 are normalized.

For the symmetry coordinates to be orthogonal it is necessary that

$$\sum_k U_{jk} U_{lk} = 0, \quad (6)$$

where j and l refer to two different symmetry coordinates. Thus for R_1 and R_2 , one has

$$(1/\sqrt{2})(0) + (1/\sqrt{2})(0) + (0)(1) = 0.$$

Similarly for R_1 and R_3 , and R_2 and R_3 , one has

$$(1/\sqrt{2})(1/\sqrt{2}) + (1/\sqrt{2})(-1/\sqrt{2}) + (0)(0) = 0$$

and

$$(0)(1/\sqrt{2}) + (0)(-1/\sqrt{2}) + (1)(0) = 0.$$

Hence R_1 , R_2 and R_3 are orthogonal.

It remains to be shown that the symmetry coordinates transform properly. For this purpose it is necessary to refer to the table of characters of the point group of the molecule. The characters of the point group C_{2v} are given in Table I. If the covering operations of the group are applied to the internal coordinates (Fig. 1), one finds that the internal coordinates are transformed as shown in Table II. Since only non-degenerate frequencies are involved, the result of applying a covering operation to a symmetry coordinate will be a transformation of the coordinate into itself or its negative, as indicated by the characters in Table I.

Consider now the symmetry coordinate R_3 . When the identity operation E is performed, each internal coordinate is transformed into itself (see Table II). Hence one may write

$$(E)R_3 = (1/\sqrt{2})\Delta d_1 - (1/\sqrt{2})\Delta d_2 = (+1)R_3.$$

However, when the operation C_2 is performed, Δd_1 is transformed into Δd_2 and Δd_2 is transformed into Δd_1 . Hence one must write

$$(C_2)R_3 = (1/\sqrt{2})\Delta d_2 - (1/\sqrt{2})\Delta d_1 = (-1)R_3,$$

TABLE II. Transformations of the internal coordinates by the covering operations of the group C_{2v} .

	E	C_2	σ_v	σ_v'
$\Delta d_1 \rightarrow$	Δd_1	Δd_2	Δd_2	Δd_1
$\Delta d_2 \rightarrow$	Δd_2	Δd_1	Δd_1	Δd_2
$\Delta\alpha \rightarrow$	$\Delta\alpha$	$\Delta\alpha$	$\Delta\alpha$	$\Delta\alpha$

which is obtained by replacing Δd_1 by Δd_2 , and Δd_2 by Δd_1 , in Eq. (3). In a similar manner, one may write

$$(\sigma_v)R_3 = (1/\sqrt{2})\Delta d_2 - (1/\sqrt{2})\Delta d_1 = (-1)R_3$$

and

$$(\sigma_v')R_3 = (1/\sqrt{2})\Delta d_1 - (1/\sqrt{2})\Delta d_2 = (+1)R_3.$$

Thus for the operations E , C_2 , σ_v and σ_v' , one gets $+1$, -1 , -1 , $+1$, respectively, as the characters for the transformations; and a comparison of these with the characters of the type B_2 vibration, shown in Table I, indicates that R_3 transforms according to the characters for the type B_2 vibration.

In a similar way one can show that R_1 and R_2 transform according to the characters for the type A_1 vibrations. Hence R_1 , R_2 and R_3 satisfy all the requirements for symmetry coordinates given in the first paragraph.

The F Matrices

To get the equations from which the vibrational frequencies can be determined, one must calculate the elements of a matrix \mathbf{F} , related to the potential energy, and a matrix \mathbf{G} , related to the kinetic energy. If one assumes harmonic motion of small amplitude for the nuclei, the expression for the potential energy V of the molecule can be written in the form⁶

$$2V = \sum f_{ik} r_i r_k, \quad (7)$$

where $f_{ik} = f_{ki}$, and i and k extend over all the internal coordinates. For the H_2O molecule, Eq. (7) becomes

$$2V = f_d [(\Delta d_1)^2 + (\Delta d_2)^2] + f_\alpha' (\Delta \alpha)^2 + 2f_{da}' (\Delta d_1 + \Delta d_2)(\Delta \alpha) + 2f_{dd} (\Delta d_1)(\Delta d_2). \quad (8)$$

Since it is customary to express all of the force constants f_{ik} in terms of the dyne per centimeter, all of the internal coordinates involving a change in angle must be multiplied by a length expressed in centimeters, usually chosen to be the equilibrium bond length of one of the bonds forming the angle. If this is done, Eq. (8) becomes

$$2V = f_d [(\Delta d_1)^2 + (\Delta d_2)^2] + f_\alpha (d \Delta \alpha)^2 + 2f_{da} (\Delta d_1 + \Delta d_2)(d \Delta \alpha) + 2f_{dd} (\Delta d_1)(\Delta d_2), \quad (9)$$

⁶ For a detailed discussion see E. T. Whittaker, *Analytical dynamics* (Cambridge, ed. 2, 1927), p. 178.

where $d [= d_1 = d_2]$ is the equilibrium value of the O-H bond distance.

But the potential energy can also be expressed in terms of the symmetry coordinates by means of the equation

$$2V = \sum F_{jl} R_j R_l, \quad (10)$$

where $F_{jl} = F_{lj}$, and j and l extend over all the symmetry coordinates.

In matrix notation,⁷ Eq. (7) and Eq. (10) become

$$2\mathbf{V} = \mathbf{r}' \mathbf{f} \mathbf{r} \quad (11)$$

and

$$2\mathbf{V} = \mathbf{R}' \mathbf{F} \mathbf{R}, \quad (12)$$

where \mathbf{r}' and \mathbf{R}' are the transposes of the \mathbf{r} and \mathbf{R} matrices, respectively. From Eqs. (11) and (12),

$$\mathbf{r}' \mathbf{f} \mathbf{r} = \mathbf{R}' \mathbf{F} \mathbf{R}. \quad (13)$$

But, as already mentioned in the preceding section,

$$\mathbf{R} = \mathbf{U} \mathbf{r}, \quad (14)$$

or

$$\mathbf{r} = \mathbf{U}^{-1} \mathbf{R}, \quad (15)$$

where \mathbf{U}^{-1} is the inverse of \mathbf{U} . But since the R_j 's are orthogonal and normalized, $\mathbf{U}^{-1} = \mathbf{U}'$. Hence Eq. (15) becomes

$$\mathbf{r} = \mathbf{U}' \mathbf{R}, \quad (16)$$

whence

$$\mathbf{r}' = (\mathbf{U}' \mathbf{R})' = \mathbf{R}' \mathbf{U}, \quad (17)$$

since the transpose of a product is equal to the product of the transposes taken in reverse order. So Eq. (13) may be written as

$$\mathbf{R}' (\mathbf{U} \mathbf{f} \mathbf{U}') \mathbf{R} = \mathbf{R}' \mathbf{F} \mathbf{R}, \quad (18)$$

whence

$$\mathbf{F} = \mathbf{U} \mathbf{f} \mathbf{U}'. \quad (19)$$

Equation (19) gives \mathbf{F} in terms of \mathbf{U} and \mathbf{f} , which is the relation necessary for obtaining the \mathbf{F} matrix.

For the H_2O molecule the \mathbf{f} matrix is

$$\begin{array}{c|ccc} & \Delta d_1 & \Delta d_2 & \Delta \alpha \\ \hline \Delta d_1 & f_d & f_{dd} & df_{da} \\ \Delta d_2 & f_{dd} & f_d & df_{da} \\ \Delta \alpha & df_{da} & df_{da} & d^2 f_\alpha \end{array}$$

⁷ For an introduction to matrices, see H. Margenau and G. M. Murphy, *The mathematics of physics and chemistry* (Van Nostrand, 1943), p. 288.

and the \mathbf{U} matrix for the type A_1 vibrations is

$$\frac{A_1}{R_1} \left| \begin{array}{ccc} \Delta d_1 & \Delta d_2 & \Delta \alpha \\ (1/\sqrt{2}) & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{array} \right|$$

Consequently, \mathbf{U}' is

$$\left(\begin{array}{cc} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1 \end{array} \right).$$

Therefore \mathbf{fU}' , which is obtained by matrix multiplication, is

$$\left(\begin{array}{cc} (1/\sqrt{2})(f_d + f_{dd}) & df_{d\alpha} \\ (1/\sqrt{2})(f_d + f_{dd}) & df_{d\alpha} \\ \sqrt{2}df_{d\alpha} & d^2f_\alpha \end{array} \right).$$

Hence, from Eq. (19), the \mathbf{F} matrix for the type A_1 vibrations is

$$\left(\begin{array}{cc} F_{11} & F_{12} \\ F_{21} & F_{22} \end{array} \right) = \left(\begin{array}{cc} f_d + f_{dd} & \sqrt{2}df_{d\alpha} \\ \sqrt{2}df_{d\alpha} & d^2f_\alpha \end{array} \right).$$

For the type B_2 vibration,

$$\mathbf{U} = (1/\sqrt{2} \quad -1/\sqrt{2} \quad 0) \quad \text{and} \quad \mathbf{U}' = \left(\begin{array}{c} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{array} \right),$$

whence

$$\mathbf{fU}' = \left(\begin{array}{c} (1/\sqrt{2})(f_d - f_{dd}) \\ (1/\sqrt{2})(f_{dd} - f_d) \\ 0 \end{array} \right)$$

and

$$\mathbf{F} = (F_{33}) = (f_d - f_{dd}),$$

a somewhat trivial case since there is only one matrix element.

The \mathbf{G} Matrices

Wilson² has shown that the elements of a modified form of the kinetic energy matrix can be obtained, when only nondegenerate vibrations are present, from the equation

$$G_{jl} = \sum_p \mu_{pg} \mathbf{S}_j^{(t)} \cdot \mathbf{S}_l^{(t)}, \quad (20)$$

where j and l refer to the symmetry coordinates used in determining the \mathbf{S} vectors; p refers to a set of equivalent atoms, a typical one of the set being t ; μ_p is the reciprocal of the mass of the typical atom t ; g_p is the number of equivalent atoms in the p th set; and the summation extends

over all the sets of equivalent atoms in the molecule. (For example, the two H atoms in the H_2O molecule form an equivalent set.)

The \mathbf{S} vectors are given by

$$\mathbf{S}_j^{(t)} = \sum_k U_{jk} \mathbf{s}_{kt}, \quad (21)$$

where j , U_{jk} and \sum_k have the same meaning as in Eq. (4). The \mathbf{s}_{kt} vectors can be expressed in terms of unit vectors directed along the chemical bonds. Referring to Fig. 2, if \mathbf{v}' is a unit vector directed along the bond from atom t to atom t' , and if r_k is the change in length of this bond from its equilibrium value, then, according to Wilson,²

$$\mathbf{s}_{kt} = -\mathbf{v}', \quad \mathbf{s}_{kt'} = \mathbf{v}'. \quad (22)$$

But if r_k is the change in the interbond angle α (Fig. 2), then the vectors for the end atoms are

$$\begin{aligned} \mathbf{s}_{kt'} &= (\cos \alpha \mathbf{v}' - \mathbf{v}'') / (d' \sin \alpha), \\ \mathbf{s}_{kt''} &= (\cos \alpha \mathbf{v}'' - \mathbf{v}') / (d'' \sin \alpha); \end{aligned} \quad (23)$$

and for the apex atom,

$$\begin{aligned} \mathbf{s}_{kt} &= [(1/d'' - \cos \alpha/d') \mathbf{v}' \\ &\quad + (1/d' - \cos \alpha/d'') \mathbf{v}''] / \sin \alpha. \end{aligned} \quad (23)$$

For the H_2O molecule, $t' = \text{H}_1$, $t = \text{O}$, $t'' = \text{H}_2$, and $d' = d'' = d_1 = d_2 = d$. Hence the \mathbf{s}_{kt} vectors for H_1 and O are

$$\begin{aligned} \mathbf{s}_{d_1\text{H}_1} &= \mathbf{v}', \\ \mathbf{s}_{\text{O}\text{H}_1} &= (\cos \alpha \mathbf{v}' - \mathbf{v}'') / (d \sin \alpha), \\ \mathbf{s}_{d_1\text{O}} &= -\mathbf{v}', \\ \mathbf{s}_{d_2\text{O}} &= -\mathbf{v}'', \\ \mathbf{s}_{\text{O}\text{O}} &= [(1 - \cos \alpha)(\mathbf{v}' + \mathbf{v}'')] / (d \sin \alpha). \end{aligned}$$

Since H_1 and H_2 are equivalent atoms, it is not necessary to calculate the \mathbf{s}_{kt} vectors for H_2 .

In order to get the $\mathbf{S}_j^{(t)}$ vectors from Eq. (21), one must first find the U_{jk} values. From Eqs. (1)-(3), it is seen that they are

$$\begin{aligned} U_{1d_1} &= 1/\sqrt{2}, & U_{2d_1} &= 0, & U_{3d_1} &= 1/\sqrt{2}, \\ U_{1d_2} &= 1/\sqrt{2}, & U_{2d_2} &= 0, & U_{3d_2} &= -1/\sqrt{2}, \\ U_{1\alpha} &= 0, & U_{2\alpha} &= 1, & U_{3\alpha} &= 0. \end{aligned}$$

Hence the $\mathbf{S}_1^{(\text{H}_1)}$ vector is

$$\begin{aligned} \mathbf{S}_1^{(\text{H}_1)} &= U_{1d_1} \mathbf{s}_{d_1\text{H}_1} + U_{1d_2} \mathbf{s}_{d_2\text{H}_1} + U_{1\alpha} \mathbf{s}_{\text{O}\text{H}_1} \\ &= (1/\sqrt{2})(\mathbf{v}') + (1/\sqrt{2})(0) \\ &\quad + (0)(\cos \alpha \mathbf{v}' - \mathbf{v}'') / (d \sin \alpha) = (1/\sqrt{2})\mathbf{v}'. \end{aligned}$$

$S_2 H_1$ being zero since a change in d_2 does not involve H_1 . Similarly,

$$S_2^{H_1} = (\cos \alpha v' - v'') / (d \sin \alpha),$$

$$S_3^{H_1} = v' / \sqrt{2},$$

$$S_1^0 = -(v' + v'') / \sqrt{2},$$

$$S_2^0 = (1 - \cos \alpha)(v' + v'') / (d \sin \alpha),$$

$$S_3^0 = (v'' - v') / \sqrt{2}.$$

The \mathbf{G} matrix for the type A_1 vibrations will contain only the four elements G_{11} , G_{12} , G_{21} and G_{22} , since only R_1 and R_2 belong to the A_1 vibrations. From Eq. (20), the value of the element G_{11} is

$$\begin{aligned} G_{11} &= \mu_H g_H S_1^{H_1} \cdot S_1^{H_1} + \mu_O g_O S_1^0 \cdot S_1^0 \\ &= \mu_H (2) (v' / \sqrt{2}) \cdot (v' / \sqrt{2}) \\ &\quad + \mu_O (1) [-(v' + v'') / \sqrt{2}] \cdot [-(v' + v'') / \sqrt{2}] \\ &= \mu_H (v' \cdot v') + \frac{1}{2} \mu_O [(v' \cdot v') \\ &\quad + (v' \cdot v'') + (v'' \cdot v') + (v'' \cdot v'')] \\ &= \mu_H + \mu_O (1 + \cos \alpha), \end{aligned}$$

since $g_H = 2$, $g_O = 1$, $v' \cdot v' = v'' \cdot v'' = 1$, and $v' \cdot v'' = v'' \cdot v' = \cos \alpha$.

The other matrix elements are obtained from Eq. (20) in a similar manner. The \mathbf{G} matrix for the type A_1 vibrations is thus

$$\begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} = \begin{pmatrix} \mu_H + \mu_O (1 + \cos \alpha) & (-\mu_O \sqrt{2} \sin \alpha) / d \\ (-\mu_O \sqrt{2} \sin \alpha) / d & 2[\mu_H + \mu_O (1 - \cos \alpha)] / d^2 \end{pmatrix}.$$

Note that $G_{12} = G_{21}$; in general, $G_{ji} = G_{lj}$.

The \mathbf{G} matrix for the type B_2 vibration will contain only the element G_{33} , since only R_3 belongs to this vibration type. Using Eq. (20), one gets

$$(G_{33}) = (\mu_H + \mu_O [1 - \cos \alpha]).$$

Calculation of the Frequencies

Having obtained the \mathbf{F} and \mathbf{G} matrices, one can get the frequencies in the following manner. For the type A_1 frequencies the secular equation in expanded form is²

$$(-\lambda)^2 + (F_{11}G_{11} + 2F_{12}G_{12} + F_{22}G_{22})(-\lambda) + \begin{vmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{vmatrix} \cdot \begin{vmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{vmatrix} = 0, \quad (24)$$

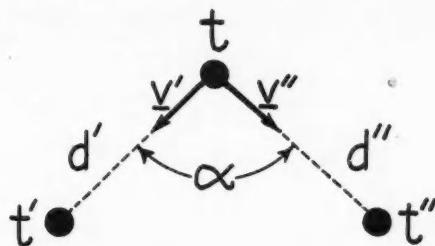


FIG. 2. Unit vectors v' and v'' directed from atom t to atoms t' and t'' , respectively, along the bonds d' and d'' , the interbond angle being α .

where the coefficient of the λ -term is the sum of the products of the corresponding one-rowed minors of the determinants of the \mathbf{F} and \mathbf{G} matrices, and the constant term is the product of the determinants of the \mathbf{F} and \mathbf{G} matrices.

At this point it is better to introduce the numerical values of the \mathbf{F} and \mathbf{G} matrix elements into Eq. (24). If the force constants, bond distances and interbond angles are known, the values of the type A_1 frequencies can be obtained from the λ 's of Eq. (24) since

$$\lambda = 4\pi^2 c^2 \nu^2, \quad (25)$$

where c (cm/sec.) is the velocity of light and ν (cm^{-1}) is the frequency. The values used here are those given by Glockler,³ namely, $\alpha = 104^\circ 31'$, $d = 0.9580 \text{ \AA}$, $f_d = 8.428 \times 10^5 \text{ dyne/cm}$,

$$\begin{aligned} f_{dd} &= -0.1051 \times 10^5 \text{ dyne/cm,} \\ f_a &= 0.7678 \times 10^5 \text{ dyne/cm,} \end{aligned}$$

and $f_{da} = 0.2521 \times 10^5 \text{ dyne/cm}$. The values used for μ_H and μ_O were

$$\mu_H = (6.022 \times 10^{23}) / 1.008 = 5.974 \times 10^{22} \text{ gm}^{-1}$$

and

$$\mu_O = (6.022 \times 10^{23}) / 16.00 = 3.764 \times 10^{22} \text{ gm}^{-1}.$$

When these quantities are used to obtain the numerical values of the \mathbf{F} and \mathbf{G} matrix elements, Eq. (24) becomes

$$\lambda^2 - (6.159 \times 10^{29} \text{ sec}^{-2}) \lambda + 5.032 \times 10^{58} \text{ sec}^{-4} = 0. \quad (26)$$

Solving this equation, one gets

$$\lambda_1 = 5.189 \times 10^{29} \text{ sec}^{-2}$$

$$\text{and} \quad \lambda_2 = 9.700 \times 10^{28} \text{ sec}^{-2}.$$

TABLE III. Comparison of the present calculated frequencies of the H_2O molecule with the values obtained by Dennison (see reference 8) using a different method.

Vibration Type	Symbol	Present value	Frequency (cm^{-1}) Dennison's value
A_1	ν_2	1653	1653.91
	ν_1	3825	3825.32
B_2	ν_3	3937	3935.59

Substituting these values of λ into Eq. (25) and using $2.998 \times 10^{10} \text{ cm/sec.}$ as the value of c , one obtains

$$\nu_1 = 3825 \text{ cm}^{-1}, \quad \nu_2 = 1653 \text{ cm}^{-1}.$$

For the type B_2 frequency the secular equation in expanded form is very simple. It is

$$\lambda_3 - F_{33}G_{33} = 0. \quad (27)$$

From Eq. (25), one gets

$$\nu_3 = (F_{33}G_{33})^{1/2}/(2\pi c). \quad (28)$$

Using the constants previously given, one finds $\nu_3 = 3937 \text{ cm}^{-1}$.

In Table III, the present values are compared with those obtained by Dennison.⁸

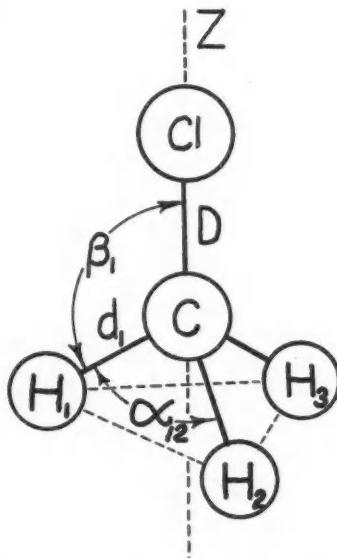


FIG. 3. Bond distances D and d_1 , interbond angles β_1 and α_{12} , and principal symmetry axis for the CH_3Cl molecule.

⁸ D. M. Dennison, *Rev. Mod. Physics* 12, 192 (1940).

Equations (24) and (28) are, of course, applicable to any nonlinear $X Y_2$ molecule, such as H_2S or SO_2 , if appropriate values of the constants³ are used.

THE CH_3Cl MOLECULE

Symmetry Coordinates

The CH_3Cl molecule has the symmetry C_{3v} and has three nondegenerate type A_1 vibrations and three doubly degenerate type E vibrations.⁵ Since there are $3(5) - 6$, or nine, vibrational frequencies, only nine symmetry coordinates are necessary. However, there are (Fig. 3) ten internal coordinates; they are ΔD , Δd_1 , Δd_2 , Δd_3 , $\Delta\beta_1$, $\Delta\beta_2$, $\Delta\beta_3$, $\Delta\alpha_{12}$, $\Delta\alpha_{23}$ and $\Delta\alpha_{31}$. Hence one of these is not independent of the others. Instead of ignoring one of the internal coordinates, which would destroy the symmetry, ten symmetry coordinates will be constructed, and then one of these will be considered as redundant. This is permissible, since the redundant coordinate can be chosen in such a way that it is identically zero and thus makes no contribution to the potential or kinetic energies.²

The symmetry coordinates chosen were the following: for the type A_1 vibrations,

$$R_1 = \Delta D, \quad (29)$$

$$R_2 = (\Delta d_1 + \Delta d_2 + \Delta d_3)/\sqrt{3}, \quad (30)$$

$$R_3 = (\Delta\alpha_{12} + \Delta\alpha_{23} + \Delta\alpha_{31}) - \Delta\beta_1 - \Delta\beta_2 - \Delta\beta_3)/\sqrt{6}, \quad (31)$$

$$R_4 = (\Delta\alpha_{12} + \Delta\alpha_{23} + \Delta\alpha_{31} + \Delta\beta_1 + \Delta\beta_2 + \Delta\beta_3)/\sqrt{6} \equiv 0 \text{ (redundant coordinate); } \quad (32)$$

and for the type E vibrations,

$$R_{1a} = (2\Delta d_1 - \Delta d_2 - \Delta d_3)/\sqrt{6}, \quad (33)$$

$$R_{1b} = (\Delta d_2 - \Delta d_3)/\sqrt{2}, \quad (34)$$

$$R_{2a} = (2\Delta\beta_1 - \Delta\beta_2 - \Delta\beta_3)/\sqrt{6}, \quad (35)$$

$$R_{2b} = (\Delta\beta_2 - \Delta\beta_3)/\sqrt{2}, \quad (36)$$

$$R_{3a} = (2\Delta\alpha_{23} - \Delta\alpha_{12} - \Delta\alpha_{31})/\sqrt{6}, \quad (37)$$

$$R_{3b} = (\Delta\alpha_{31} - \Delta\alpha_{12})/\sqrt{2}. \quad (38)$$

The coordinate R_4 was chosen as the redundant coordinate, since it is clear that the sum of all the changes in the interbond angles must be zero. The remaining coordinates were made orthogonal to R_4 and all were normalized, using Eqs. (5) and

(6). Since each of the three type *E* vibrations is doubly degenerate, it is necessary to have two symmetry coordinates, designated by the subscripts *a* and *b*, for each of these vibrations.

The transformations of the internal coordinates under the covering operations of the group C_{3v} are shown in Table IV. The characters for the group C_{3v} are given in Table V. Proceeding in the same manner as for the H_2O molecule, using Table IV, one can show that the A_1 symmetry coordinates R_1 , R_2 , R_3 and R_4 transform according to the characters for the type A_1 vibrations, given in Table V.

When a covering operation is applied to one of a pair of the *E* coordinates, one obtains not this coordinate or its negative, as was the case for the coordinates associated with the non-degenerate frequencies, but rather a linear combination of the two coordinates forming the pair. For example, when the operation C_3^+ is applied to the pair R_{1a} , R_{1b} , one has

$$(C_3^+)R_{1a} = (2\Delta d_2 - \Delta d_3 - \Delta d_1)/(6)^{\frac{1}{2}} = A R_{1a} + B R_{1b}$$

and

$$(C_3^+)R_{1b} = (\Delta d_3 - \Delta d_1)/\sqrt{2} = A' R_{1a} + B' R_{1b},$$

where A , B , A' and B' are constants. Substituting for R_{1a} and R_{1b} from Eqs. (33) and (34), one gets $(2\Delta d_2 - \Delta d_3 - \Delta d_1)/(6)^{\frac{1}{2}}$

$$= A(2\Delta d_1 - \Delta d_2 - \Delta d_3)/(6)^{\frac{1}{2}} + B(\Delta d_2 - \Delta d_3)/\sqrt{2}$$

and

$$(\Delta d_3 - \Delta d_1)/\sqrt{2} = A'(2\Delta d_1 - \Delta d_2 - \Delta d_3)/(6)^{\frac{1}{2}} + B'(\Delta d_2 - \Delta d_3)/\sqrt{2}.$$

Equating coefficients, one finds that

$$\begin{aligned} A &= -\frac{1}{2}, & B &= \sqrt{3}/2, \\ A' &= -\sqrt{3}/2, & B' &= -\frac{1}{2}. \end{aligned}$$

If one now forms the matrix,

$$\begin{pmatrix} A & B \\ A' & B' \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \sqrt{3}/2 \\ -\sqrt{3}/2 & -\frac{1}{2} \end{pmatrix},$$

the sum of the elements along the principal diagonal, $-\frac{1}{2} - \frac{1}{2} = -1$, gives the character under C_3 in Table V for the type *E* vibrations. The same character would have been obtained if C_3^- had been used.

In a similar manner, when the operation E is

TABLE IV. Transformations of the internal coordinates of the CH_3Cl molecule under the covering operations of the group C_{3v} .*

	<i>E</i>	C_3^+	C_3^-	σ_{11}	σ_{12}	σ_{13}
$\Delta D \rightarrow$	ΔD					
$\Delta d_{1 \rightarrow}$	Δd_1	Δd_2	Δd_3	Δd_1	Δd_3	Δd_2
$\Delta d_{2 \rightarrow}$	Δd_2	Δd_3	Δd_1	Δd_3	Δd_2	Δd_1
$\Delta d_{3 \rightarrow}$	Δd_3	Δd_1	Δd_2	Δd_2	Δd_1	Δd_3
$\Delta \alpha_{12 \rightarrow}$	$\Delta \alpha_{12}$	$\Delta \alpha_{23}$	$\Delta \alpha_{31}$	$\Delta \alpha_{31}$	$\Delta \alpha_{23}$	$\Delta \alpha_{12}$
$\Delta \alpha_{23 \rightarrow}$	$\Delta \alpha_{23}$	$\Delta \alpha_{31}$	$\Delta \alpha_{12}$	$\Delta \alpha_{23}$	$\Delta \alpha_{12}$	$\Delta \alpha_{31}$
$\Delta \alpha_{31 \rightarrow}$	$\Delta \alpha_{31}$	$\Delta \alpha_{12}$	$\Delta \alpha_{23}$	$\Delta \alpha_{12}$	$\Delta \alpha_{31}$	$\Delta \alpha_{23}$
$\Delta \beta_{1 \rightarrow}$	$\Delta \beta_1$	$\Delta \beta_2$	$\Delta \beta_3$	$\Delta \beta_1$	$\Delta \beta_3$	$\Delta \beta_2$
$\Delta \beta_{2 \rightarrow}$	$\Delta \beta_2$	$\Delta \beta_3$	$\Delta \beta_1$	$\Delta \beta_2$	$\Delta \beta_3$	$\Delta \beta_1$
$\Delta \beta_{3 \rightarrow}$	$\Delta \beta_3$	$\Delta \beta_1$	$\Delta \beta_2$	$\Delta \beta_3$	$\Delta \beta_1$	$\Delta \beta_2$

* C_3^+ = rotation by $+2\pi/3$, C_3^- = rotation by $-2\pi/3$ about the *Z* axis; σ_{11} , σ_{12} and σ_{13} are reflection planes through the *Z* axis and through H_1 , H_2 and H_3 , respectively

TABLE V. Characters for the group C_{3v} .

C_{3v}	<i>E</i>	$2C_3$	$3\sigma_g$
A_1	1	1	1
A_2	1	1	-1
E	2	-1	0

applied to R_{1a} and R_{1b} one gets the matrix

$$\begin{pmatrix} A & B \\ A' & B' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

from which the character, 2, is obtained, in agreement with Table V. For the operation σ_{11} , the corresponding matrix is

$$\begin{pmatrix} A & B \\ A' & B' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

the sum of the elements along the principal diagonal being zero, in agreement with Table V. Had σ_{12} or σ_{13} been used, the same character would have been obtained.

One has proved, therefore, that the *E* coordinate pair, R_{1a} , R_{1b} , transforms according to the characters of the type *E* vibrations. In a similar manner, one can show that the coordinate pairs R_{2a} , R_{2b} and R_{3a} , R_{3b} also transform in this way.

Hence it has been demonstrated that the symmetry coordinates satisfy all the requirements given in the discussion of the symmetry coordinates for the H_2O molecule.

The F Matrices

For the CH_3Cl molecule it is not possible to use the most general quadratic expression for the

potential energy, Eq. (7), as was done for the H_2O molecule, since more force constants would be involved than could be determined from the observed frequencies. Consequently, one is forced to neglect some of the terms which, upon the basis of experience, one suspects would make only small contributions to the potential energy. Crawford and Brinkley⁹ have given a rule that is helpful in deciding which terms to neglect.

The potential function used for the CH_3Cl molecule was

$$\begin{aligned} 2V = & f_d[(\Delta d_1)^2 + (\Delta d_2)^2 + (\Delta d_3)^2] + f_\alpha[(d\Delta\alpha_{12})^2 + (d\Delta\alpha_{23})^2 + (d\Delta\alpha_{31})^2] + 2f_{\alpha\beta}[(d\Delta\beta_1)(d\Delta\alpha_{12} + d\Delta\alpha_{31}) \\ & + (d\Delta\beta_2)(d\Delta\alpha_{12} + d\Delta\alpha_{23}) + (d\Delta\beta_3)(d\Delta\alpha_{23} + d\Delta\alpha_{31})] + f_\beta[(d\Delta\beta_1)^2 + (d\Delta\beta_2)^2 + (d\Delta\beta_3)^2] \\ & + f_D(\Delta D)^2 + 2f_{D\alpha}[(\Delta D)(d\Delta\alpha_{12} + d\Delta\alpha_{23} + d\Delta\alpha_{31})] + 2f_{D\beta}[(\Delta D)(d\Delta\beta_1 + d\Delta\beta_2 + d\Delta\beta_3)]. \quad (39) \end{aligned}$$

The corresponding f matrix is

$$\begin{array}{|c|ccccccccccccc|} \hline & \Delta D & \Delta d_1 & \Delta d_2 & \Delta d_3 & \Delta\beta_1 & \Delta\beta_2 & \Delta\beta_3 & \Delta\alpha_{12} & \Delta\alpha_{23} & \Delta\alpha_{31} \\ \hline \Delta D & f_D & 0 & 0 & 0 & df_{D\beta} & df_{D\beta} & df_{D\beta} & df_{D\alpha} & df_{D\alpha} & df_{D\alpha} \\ \Delta d_1 & 0 & f_d & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \Delta d_2 & 0 & 0 & f_d & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \Delta d_3 & 0 & 0 & 0 & f_d & 0 & 0 & 0 & 0 & 0 & 0 \\ \Delta\beta_1 & df_{D\beta} & 0 & 0 & 0 & d^2f_{\beta} & 0 & 0 & d^2f_{\alpha\beta} & 0 & d^2f_{\alpha\beta} \\ \Delta\beta_2 & df_{D\beta} & 0 & 0 & 0 & 0 & d^2f_{\beta} & 0 & d^2f_{\alpha\beta} & d^2f_{\alpha\beta} & 0 \\ \Delta\beta_3 & df_{D\beta} & 0 & 0 & 0 & 0 & 0 & d^2f_{\beta} & 0 & d^2f_{\alpha\beta} & d^2f_{\alpha\beta} \\ \Delta\alpha_{12} & df_{D\alpha} & 0 & 0 & 0 & d^2f_{\alpha\beta} & d^2f_{\alpha\beta} & 0 & d^2f_{\alpha} & 0 & 0 \\ \Delta\alpha_{23} & df_{D\alpha} & 0 & 0 & 0 & 0 & d^2f_{\alpha\beta} & d^2f_{\alpha\beta} & 0 & d^2f_{\alpha} & 0 \\ \Delta\alpha_{31} & df_{D\alpha} & 0 & 0 & 0 & d^2f_{\alpha\beta} & 0 & d^2f_{\alpha\beta} & 0 & 0 & d^2f_{\alpha} \\ \hline \end{array}.$$

For the type A_1 vibrations the \mathbf{U} and \mathbf{U}' matrices are

$$\mathbf{U} = R_1 \begin{pmatrix} \Delta D & \Delta d_1 & \Delta d_2 & \Delta d_3 & \Delta\beta_1 & \Delta\beta_2 & \Delta\beta_3 & \Delta\alpha_{12} & \Delta\alpha_{23} & \Delta\alpha_{31} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix}$$

and

$$\mathbf{U}' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & -1/\sqrt{6} \\ 0 & 0 & -1/\sqrt{6} \\ 0 & 0 & -1/\sqrt{6} \\ 0 & 0 & 1/\sqrt{6} \\ 0 & 0 & 1/\sqrt{6} \\ 0 & 0 & 1/\sqrt{6} \end{pmatrix}.$$

⁹ B. L. Crawford, Jr. and S. R. Brinkley, Jr., *J. Chem. Physics* 9, 71 (1941).

From Eq. (19), the \mathbf{F} matrix for the type A_1 vibrations is found to be

$$\begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix} = \begin{pmatrix} f_D & 0 & [d\sqrt{3}(f_{D\alpha}-f_{D\beta})]/2 \\ 0 & f_d & 0 \\ [d\sqrt{3}(f_{D\alpha}-f_{D\beta})]/2 & 0 & [d^2(f_\alpha+f_\beta-4f_{\alpha\beta})]/2 \end{pmatrix}.$$

Using R_{1b} , R_{2b} and R_{3b} , the \mathbf{U} and \mathbf{U}' matrices for the type E vibrations are found to be

$$\mathbf{U} = \begin{pmatrix} 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}$$

and

$$\mathbf{U}' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1/\sqrt{2} & 0 & 0 \\ -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 0 \\ 0 & 0 & -1/\sqrt{2} \\ 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} \end{pmatrix}.$$

From Eq. (19), the \mathbf{F} matrix for the type E vibrations is found to be

$$\begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix} = \begin{pmatrix} f_d & 0 & 0 \\ 0 & d^2f_\beta & -d^2f_{\alpha\beta} \\ 0 & -d^2f_{\alpha\beta} & d^2f_\alpha \end{pmatrix}.$$

The same \mathbf{F} matrix could have been obtained in the same way by use of R_{1a} , R_{2a} and R_{3a} .

The G Matrices

The elements of the \mathbf{G} matrix for the non-degenerate type A_1 vibrations can be obtained from Eq. (20). However, for the doubly degenerate type E vibrations, the corresponding equation for the elements of the \mathbf{G} matrix is²

$$G_{jl} = \frac{1}{2} \sum_p \mu_p g_p (\mathbf{S}_{ja}^{(t)} \cdot \mathbf{S}_{la}^{(t)} + \mathbf{S}_{jb}^{(t)} \cdot \mathbf{S}_{lb}^{(t)}), \quad (40)$$

where $\mathbf{S}_{ja}^{(t)}$ and $\mathbf{S}_{jb}^{(t)}$ are the vectors, for the atom t , obtained from the U 's appearing in the symmetry coordinate pair R_{ja} , R_{jb} , and the other symbols have the same meaning as in Eq. (20).

Using Eqs. (22) and (23) and the unit vectors, bond distances, and interbond angles indicated in Fig. 4, and assuming tetrahedral angles ($\alpha = \beta = 109^\circ 28'$), the \mathbf{s}_{kt} vectors for the typical atoms are found to be: for the H_1 atom, letting

$$\epsilon = 1/d \text{ and } \tau = 1/D,$$

$$\mathbf{s}_{d1H_1} = \mathbf{v}_1,$$

$$\mathbf{s}_{\alpha 12H_1} = -\frac{3}{4}\sqrt{2}\epsilon(\mathbf{v}_1/3 + \mathbf{v}_2),$$

$$\mathbf{s}_{\alpha 31H_1} = -\frac{3}{4}\sqrt{2}\epsilon(\mathbf{v}_1/3 + \mathbf{v}_3),$$

$$\mathbf{s}_{\beta 1H_1} = -\frac{3}{4}\sqrt{2}\epsilon(\mathbf{v}_1/3 + \mathbf{v}),$$

$$\mathbf{s}_{\alpha 23H_1} = \mathbf{s}_{d2H_1} = \mathbf{s}_{d3H_1} = \mathbf{s}_{DH_1} = \mathbf{s}_{\beta 2H_1} = \mathbf{s}_{\beta 3H_1} = 0;$$

for the C atom

$$\mathbf{s}_{DC} = -\mathbf{v},$$

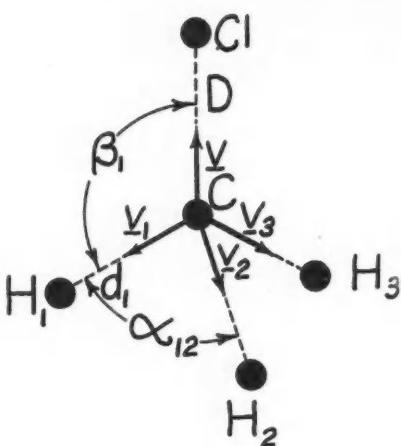
$$\mathbf{s}_{d1C} = -\mathbf{v}_1,$$

$$\mathbf{s}_{d2C} = -\mathbf{v}_2,$$

$$\mathbf{s}_{d3C} = -\mathbf{v}_3,$$

$$\mathbf{s}_{\alpha 12C} = \epsilon\sqrt{2}(\mathbf{v}_1 + \mathbf{v}_2),$$

$$\mathbf{s}_{\alpha 23C} = \epsilon\sqrt{2}(\mathbf{v}_2 + \mathbf{v}_3),$$

FIG. 4. Unit vectors for the CH_3Cl molecule.

$$\begin{aligned} S_{\alpha_31}\text{C} &= \epsilon\sqrt{2}(\mathbf{v}_3 + \mathbf{v}_1), \\ S_{\beta_1}\text{C} &= \frac{3}{4}\sqrt{2}[(\epsilon + \tau/3)\mathbf{v} + (\tau + \epsilon/3)\mathbf{v}_1], \\ S_{\beta_2}\text{C} &= \frac{3}{4}\sqrt{2}[(\epsilon + \tau/3)\mathbf{v} + (\tau + \epsilon/3)\mathbf{v}_2], \\ S_{\beta_3}\text{C} &= \frac{3}{4}\sqrt{2}[(\epsilon + \tau/3)\mathbf{v} + (\tau + \epsilon/3)\mathbf{v}_3]; \end{aligned}$$

for the Cl atom

$$\begin{aligned} S_{DC1} &= \mathbf{v}, \\ S_{\beta_1}\text{Cl} &= -\frac{3}{4}\sqrt{2}\tau(\mathbf{v}/3 + \mathbf{v}_1), \\ S_{\beta_2}\text{Cl} &= -\frac{3}{4}\sqrt{2}\tau(\mathbf{v}/3 + \mathbf{v}_2), \\ S_{\beta_3}\text{Cl} &= -\frac{3}{4}\sqrt{2}\tau(\mathbf{v}/3 + \mathbf{v}_3), \\ S_{d_1}\text{Cl} = S_{d_2}\text{Cl} = S_{d_3}\text{Cl} = S_{\alpha_{12}}\text{Cl} = S_{\alpha_{23}}\text{Cl} = S_{\alpha_{31}}\text{Cl} &= 0. \end{aligned}$$

From Eqs. (29) to (38), the U 's are found to be: for the A_1 symmetry coordinates,

$$\begin{aligned} U_{1D} &= 1, \text{ all other } U_{1k} = 0, \\ U_{2d_1} = U_{2d_2} = U_{2d_3} &= 1/\sqrt{3}, \text{ all other } U_{2k} = 0, \\ U_{3\beta_1} = U_{3\beta_2} = U_{3\beta_3} &= -1/\sqrt{6}, \\ U_{3\alpha_{12}} = U_{3\alpha_{23}} = U_{3\alpha_{31}} &= 1/\sqrt{6}, \text{ and all other } U_{3k} = 0; \\ \text{for the } E \text{ symmetry coordinates,} \\ U_{1ad_1} &= 2/\sqrt{6}, \quad U_{1ad_2} = U_{1ad_3} = -1/\sqrt{6}, \\ &\quad \text{all other } U_{1ak} = 0, \\ U_{2a\beta_1} &= 2/\sqrt{6}, \quad U_{2a\beta_2} = U_{2a\beta_3} = -1/\sqrt{6}, \\ &\quad \text{all other } U_{2ak} = 0, \\ U_{3a\alpha_{23}} &= 2/\sqrt{6}, \quad U_{3a\alpha_{12}} = U_{3a\alpha_{31}} = -1/\sqrt{6}, \\ &\quad \text{all other } U_{3ak} = 0, \end{aligned}$$

$$\begin{aligned} U_{1bd_2} &= 1/\sqrt{2}, \quad U_{1bd_3} = -1/\sqrt{2}, \text{ all other } U_{1bk} = 0, \\ U_{2bb_2} &= 1/\sqrt{2}, \quad U_{2bb_3} = -1/\sqrt{2}, \text{ all other } U_{2bk} = 0, \\ U_{3ba_31} &= 1/\sqrt{2}, \quad U_{3ba_{12}} = -1/\sqrt{2}, \\ &\quad \text{and all other } U_{3bk} = 0. \end{aligned}$$

Now using Eq. (21), one finds the \mathbf{S} vectors to be: for the A_1 vibrations,

$$\begin{aligned} \mathbf{S}_1^{H_1} &= 0, \\ \mathbf{S}_2^{H_1} &= (1/\sqrt{3})\mathbf{v}_1, \\ \mathbf{S}_3^{H_1} &= (-\epsilon\sqrt{3}/12)(\mathbf{v}_1 + 3\mathbf{v}_2 + 3\mathbf{v}_3 - 3\mathbf{v}), \\ \mathbf{S}_1^C &= -\mathbf{v}, \\ \mathbf{S}_2^C &= (-1/\sqrt{3})(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3), \\ \mathbf{S}_3^C &= (7\epsilon\sqrt{3}/12 - \sqrt{3}\tau/4)(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3) \\ &\quad - (\sqrt{3}/4)(3\epsilon + \tau)\mathbf{v}, \\ \mathbf{S}_1^{Cl} &= \mathbf{v}, \\ \mathbf{S}_2^{Cl} &= 0, \\ \mathbf{S}_3^{Cl} &= (\tau\sqrt{3}/4)(\mathbf{v} + \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3); \end{aligned}$$

for the E vibrations,

$$\begin{aligned} \mathbf{S}_{1a}^{H_1} &= (2/\sqrt{6})\mathbf{v}_1, \\ \mathbf{S}_{2a}^{H_1} &= (-\epsilon\sqrt{3}/2)(\mathbf{v}_1/3 + \mathbf{v}), \\ \mathbf{S}_{3a}^{H_1} &= (\epsilon\sqrt{3}/4)(2\mathbf{v}_1/3 + \mathbf{v}_2 + \mathbf{v}_3), \\ \mathbf{S}_{1a}^C &= (1/\sqrt{6})(-2\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3), \\ \mathbf{S}_{2a}^C &= (\sqrt{3}/4)(\tau + \epsilon/3)(2\mathbf{v}_1 - \mathbf{v}_2 - \mathbf{v}_3), \\ \mathbf{S}_{3a}^C &= (\epsilon\sqrt{3}/3)(-2\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3), \\ \mathbf{S}_{1a}^{Cl} &= 0, \\ \mathbf{S}_{2a}^{Cl} &= (\tau\sqrt{3}/4)(-2\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3), \\ \mathbf{S}_{3a}^{Cl} &= 0, \\ \mathbf{S}_{1b}^{H_1} &= 0, \\ \mathbf{S}_{2b}^{H_1} &= 0, \\ \mathbf{S}_{3b}^{H_1} &= \frac{3}{4}\epsilon(\mathbf{v}_2 - \mathbf{v}_3), \\ \mathbf{S}_{1b}^C &= (1/\sqrt{2})(\mathbf{v}_3 - \mathbf{v}_2), \\ \mathbf{S}_{2b}^C &= \frac{3}{4}(\tau + \epsilon/3)(\mathbf{v}_2 - \mathbf{v}_3), \\ \mathbf{S}_{3b}^C &= \epsilon(\mathbf{v}_3 - \mathbf{v}_2), \\ \mathbf{S}_{1b}^{Cl} &= 0, \\ \mathbf{S}_{2b}^{Cl} &= \frac{3}{4}\tau(\mathbf{v}_3 - \mathbf{v}_2), \\ \mathbf{S}_{3b}^{Cl} &= 0. \end{aligned}$$

The \mathbf{G} matrix for the type A_1 vibrations is obtained by use of Eq. (20), just as for the H_2O molecule; it is

$$\begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{bmatrix} = \begin{bmatrix} \mu_{\text{Cl}} + \mu_{\text{C}} & (-\sqrt{3}/3)\mu_{\text{C}} & (4\epsilon\sqrt{3}/3)\mu_{\text{C}} \\ (-\sqrt{3}/3)\mu_{\text{C}} & \mu_{\text{H}} + \mu_{\text{C}}/3 & (-4\epsilon/3)\mu_{\text{C}} \\ (4\epsilon\sqrt{3}/3)\mu_{\text{C}} & (-4\epsilon/3)\mu_{\text{C}} & 2\epsilon^2\mu_{\text{H}} + (16/3)\epsilon^2\mu_{\text{C}} \end{bmatrix}.$$

The **G** matrix elements for the type *E* vibrations are obtained by use of Eq. (40). For example, the G_{23} element would be

$$G_{23} = \frac{1}{2} [\mu_{\text{H}}(3)(\mathbf{S}_{2a}^{\text{H}1} \cdot \mathbf{S}_{3a}^{\text{H}1} + \mathbf{S}_{2b}^{\text{H}1} \cdot \mathbf{S}_{3b}^{\text{H}1}) + \mu_{\text{C}}(1)(\mathbf{S}_{2a}^{\text{C}} \cdot \mathbf{S}_{3a}^{\text{C}} + \mathbf{S}_{2b}^{\text{C}} \cdot \mathbf{S}_{3b}^{\text{C}}) + \mu_{\text{Cl}}(1)(\mathbf{S}_{2a}^{\text{Cl}} \cdot \mathbf{S}_{3a}^{\text{Cl}} + \mathbf{S}_{2b}^{\text{Cl}} \cdot \mathbf{S}_{3b}^{\text{Cl}})]. \quad (41)$$

The dot products are

$$\begin{aligned} \mathbf{S}_{2a}^{\text{H}1} \cdot \mathbf{S}_{3a}^{\text{H}1} &= \epsilon^2/3, & \mathbf{S}_{2a}^{\text{C}} \cdot \mathbf{S}_{3a}^{\text{C}} &= -2\epsilon(\tau + \epsilon/3), & \mathbf{S}_{2a}^{\text{Cl}} \cdot \mathbf{S}_{3a}^{\text{Cl}} &= 0, \\ \mathbf{S}_{2b}^{\text{H}1} \cdot \mathbf{S}_{3b}^{\text{H}1} &= 0, & \mathbf{S}_{2b}^{\text{C}} \cdot \mathbf{S}_{3b}^{\text{C}} &= -2\epsilon(\tau + \epsilon/3), & \mathbf{S}_{2b}^{\text{Cl}} \cdot \mathbf{S}_{3b}^{\text{Cl}} &= 0. \end{aligned}$$

Hence, $G_{23} = \frac{1}{2}\epsilon^2\mu_{\text{H}} - 2\epsilon(\tau + \epsilon/3)\mu_{\text{C}}$. The complete matrix for the type *E* vibrations is

$$\begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{bmatrix} = \begin{bmatrix} \mu_{\text{H}} + (4/3)\mu_{\text{C}} & -\sqrt{2}(\tau + \epsilon/3)\mu_{\text{C}} & (4\epsilon\sqrt{2}/3)\mu_{\text{C}} \\ -\sqrt{2}\mu_{\text{C}}(\tau + \epsilon/3) & (3\tau^2/2)\mu_{\text{C}} + \epsilon^2\mu_{\text{H}} + (3/2)(\tau + \epsilon/3)^2\mu_{\text{C}} & (\epsilon^2/2)\mu_{\text{H}} - 2\epsilon(\tau + \epsilon/3)\mu_{\text{C}} \\ (4\epsilon\sqrt{2}/3)\mu_{\text{C}} & (\epsilon^2/2)\mu_{\text{H}} - 2\epsilon(\tau + \epsilon/3)\mu_{\text{C}} & (5\epsilon^2/2)\mu_{\text{H}} + (8\epsilon^2/3)\mu_{\text{C}} \end{bmatrix}.$$

Calculation of the Frequencies

The secular equation in expanded form² for the type *A*₁ frequencies obtained from the **F** and **G** matrices for the *A*₁ vibrations is

$$\begin{aligned} &(-\lambda)^3 + (F_{11}G_{11} + F_{22}G_{22} + F_{33}G_{33} + 2F_{12}G_{12} + 2F_{13}G_{13} + 2F_{23}G_{23})(-\lambda)^2 \\ &+ \left\{ \begin{vmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{vmatrix} \cdot \begin{vmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{vmatrix} + \begin{vmatrix} F_{12} & F_{13} \\ F_{22} & F_{23} \end{vmatrix} \cdot \begin{vmatrix} G_{12} & G_{13} \\ G_{22} & G_{23} \end{vmatrix} + \begin{vmatrix} F_{11} & F_{13} \\ F_{21} & F_{23} \end{vmatrix} \cdot \begin{vmatrix} G_{11} & G_{13} \\ G_{21} & G_{23} \end{vmatrix} \right. \\ &+ \begin{vmatrix} F_{11} & F_{12} \\ F_{31} & F_{32} \end{vmatrix} \cdot \begin{vmatrix} G_{11} & G_{12} \\ G_{31} & G_{32} \end{vmatrix} + \begin{vmatrix} F_{12} & F_{13} \\ F_{32} & F_{33} \end{vmatrix} \cdot \begin{vmatrix} G_{12} & G_{13} \\ G_{32} & G_{33} \end{vmatrix} + \begin{vmatrix} F_{11} & F_{13} \\ F_{31} & F_{33} \end{vmatrix} \cdot \begin{vmatrix} G_{11} & G_{13} \\ G_{31} & G_{33} \end{vmatrix} \\ &\left. + \begin{vmatrix} F_{21} & F_{22} \\ F_{31} & F_{32} \end{vmatrix} \cdot \begin{vmatrix} G_{21} & G_{22} \\ G_{31} & G_{32} \end{vmatrix} + \begin{vmatrix} F_{22} & F_{23} \\ F_{32} & F_{33} \end{vmatrix} \cdot \begin{vmatrix} G_{22} & G_{23} \\ G_{32} & G_{33} \end{vmatrix} + \begin{vmatrix} F_{21} & F_{23} \\ F_{31} & F_{33} \end{vmatrix} \cdot \begin{vmatrix} G_{21} & G_{23} \\ G_{31} & G_{33} \end{vmatrix} \right\} (-\lambda) \\ &+ \begin{vmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{vmatrix} \cdot \begin{vmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{vmatrix} = 0. \quad (42) \end{aligned}$$

It will be observed that the coefficient of $(-\lambda)^2$ is the sum of the products of all the corresponding one-rowed minors of the determinants of the **F** and **G** matrices; the coefficient of $(-\lambda)$ is the sum of the products of all the corresponding two-rowed minors; and the constant term is, as always, the product of the determinants of the **F** and **G** matrices.

The numerical values of the matrix elements can now be calculated, using the following values of the force constants,⁹ bond distances⁹ and reciprocals of masses:

$$f_D = 3.64 \times 10^5 \text{ dyne/cm},$$

$$f_d = 4.790 \times 10^5 \text{ dyne/cm},$$

$$f_a = 0.46 \times 10^5 \text{ dyne/cm},$$

$$f_\beta = 0.58 \times 10^5 \text{ dyne/cm},$$

$$f_{ab} = 0.01 \times 10^6 \text{ dyne/cm},$$

$$f_{D\alpha} - f_{D\beta} = -0.49 \times 10^5 \text{ dyne/cm},$$

$$d = 1.093A = 1/\epsilon,$$

$$D = 1.66A = 1/\tau,$$

$$\mu_{\text{Cl}} = 1.698 \times 10^{22} \text{ gm}^{-1},$$

TABLE VI. Comparison of the calculated and observed values of the vibrational frequencies of the CH_3Cl molecule.

Type	Calculated (cm^{-1})	Crawford and Brinkley (see reference 9)	Observed (see reference 9) (cm^{-1})
	Present values		
A_1	758	732*	732
	1371	1355*	1355
	2883	2902	2920
E	1012	1019	1020
	1468	1478	1460
	3001	3021	3047

* Used in obtaining the force constants.

$$\mu_C = 5.018 \times 10^{22} \text{ gm}^{-1},$$

$$\mu_H = 5.974 \times 10^{23} \text{ gm}^{-1}.$$

Upon substitution of the numerical values, Eq. (42) becomes

$$\lambda^3 - (3.819 \times 10^{29} \text{ sec}^{-2})\lambda^2 + (2.703 \times 10^{58} \text{ sec}^{-4})\lambda - 4.006 \times 10^{86} \text{ sec}^{-6} = 0. \quad (43)$$

Solution of this equation gives

$$\lambda_1 = 2.948 \times 10^{29} \text{ sec}^{-2},$$

$$\lambda^2 = 0.2037 \times 10^{29} \text{ sec}^{-2},$$

$$\lambda_3 = 0.6668 \times 10^{29} \text{ sec}^{-2},$$

and the corresponding values of ν from Eq. (25) are

$$\nu_1 = 2883 \text{ cm}^{-1}, \quad \nu_2 = 758 \text{ cm}^{-1}, \quad \nu_3 = 1371 \text{ cm}^{-1}.$$

The secular equation in expanded form for the type E frequencies is identical with Eq. (42), except that now the F and G matrix elements obtained for the E vibrations must be used. Evaluation and substitution of these, using the previously listed constants, gives

$$\lambda^3 - (4.322 \times 10^{29} \text{ sec}^{-2})\lambda^2 + (3.881 \times 10^{58} \text{ sec}^{-4})\lambda - 8.875 \times 10^{86} \text{ sec}^{-6} = 0. \quad (44)$$

Solution of this equation gives

$$\lambda_4 = 3.195 \times 10^{29} \text{ sec}^{-2},$$

$$\lambda_5 = 0.3634 \times 10^{29} \text{ sec}^{-2},$$

$$\lambda_6 = 0.7643 \times 10^{29} \text{ sec}^{-2};$$

the corresponding values of ν are

$$\nu_4 = 3001 \text{ cm}^{-1}, \quad \nu_5 = 1012 \text{ cm}^{-1}, \quad \nu_6 = 1468 \text{ cm}^{-1}.$$

The present calculated values of the A_1 and E frequencies are compared with the values calculated by Crawford and Brinkley⁹ and with the observed values in Table VI.

It should be noted that Eq. (42) and the F and G matrices are valid for any XY_3Z molecule whose symmetry is C_{3v} , such as CD_3Cl , CH_3Br , CCl_3H , if the angles are tetrahedral and if appropriate values of the constants are used.

THE CH_4 AND CD_4 MOLECULES

Symmetry Coordinates

The CH_4 molecule has the symmetry T_d and therefore has one nondegenerate type A_1 vibration, one doubly degenerate type E vibration, and two triply degenerate type T_2 vibrations.⁵ Although there are the same number of vibrational frequencies for this molecule as for CH_3Cl , the greater symmetry allows triply degenerate vibrations to occur. Again there are ten internal coordinates, but now all the interbond angles as well as all the bond distances are equivalent; the internal coordinates are (see Fig. 5) Δd_1 , Δd_2 , Δd_3 , Δd_4 , $\Delta\alpha_{12}$, $\Delta\alpha_{23}$, $\Delta\alpha_{31}$, $\Delta\alpha_{14}$, $\Delta\alpha_{24}$ and $\Delta\alpha_{34}$.

As before, ten symmetry coordinates, one of which is redundant, must be constructed from these internal coordinates. The symmetry coordinates used were: for the type A_1 vibrations,

$$R_1 = \frac{1}{2}(\Delta d_1 + \Delta d_2 + \Delta d_3 + \Delta d_4), \quad (45)$$

$$R_2 = (1/\sqrt{6})(\Delta\alpha_{12} + \Delta\alpha_{23} + \Delta\alpha_{31} + \Delta\alpha_{14} + \Delta\alpha_{24} + \Delta\alpha_{34}) \equiv 0 \text{ (redundant coordinate); } \quad (46)$$

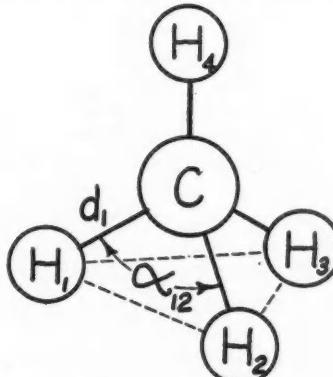


FIG. 5. The CH_4 molecule.

for the type E vibrations,

$$\begin{aligned} R_{1a} &= (1/\sqrt{12})(2\Delta\alpha_{12} - \Delta\alpha_{23} - \Delta\alpha_{31} \\ &\quad - \Delta\alpha_{14} - \Delta\alpha_{24} + 2\Delta\alpha_{34}), \end{aligned} \quad (47)$$

$$R_{1b} = \frac{1}{2}(\Delta\alpha_{14} - \Delta\alpha_{31} + \Delta\alpha_{23} - \Delta\alpha_{24}); \quad (48)$$

and for the type T_2 vibrations,

$$\begin{aligned} R_{1a} &= (1/\sqrt{12})(2\Delta\alpha_{12} - \Delta\alpha_{23} - \Delta\alpha_{31} + \Delta\alpha_{14} \\ &\quad + \Delta\alpha_{24} - 2\Delta\alpha_{34}), \end{aligned} \quad (49)$$

$$\begin{aligned} R_{1b} &= (1/\sqrt{6})(\Delta\alpha_{12} + \Delta\alpha_{23} + \Delta\alpha_{31} - \Delta\alpha_{14} \\ &\quad - \Delta\alpha_{24} - \Delta\alpha_{34}), \end{aligned} \quad (50)$$

$$R_{1c} = \frac{1}{2}(\Delta\alpha_{23} - \Delta\alpha_{31} - \Delta\alpha_{14} + \Delta\alpha_{24}), \quad (51)$$

$$R_{2a} = (1/\sqrt{6})(\Delta d_1 + \Delta d_2 - 2\Delta d_3), \quad (52)$$

$$R_{2b} = (1/\sqrt{12})(\Delta d_1 + \Delta d_2 + \Delta d_3 - 3\Delta d_4), \quad (53)$$

$$R_{2c} = (1/\sqrt{2})(\Delta d_2 - \Delta d_1). \quad (54)$$

The symmetry coordinates for the A_1 and E vibrations can be shown to transform properly by the same method that was used for the CH_3Cl molecule. A similar method is used for the type T_2 vibrations, except that when a covering operation is applied to one of a set of three T_2 symmetry coordinates (say R_{1a} , R_{1b} , R_{1c}), the result is a linear combination of the three symmetry coordinates forming the set. For example, for the covering operation σ_d , one would have

$$\begin{aligned} 2V &= f_d[(\Delta d_1)^2 + (\Delta d_2)^2 + (\Delta d_3)^2 + (\Delta d_4)^2] \\ &\quad + f_a[(d\Delta\alpha_{12})^2 + (d\Delta\alpha_{23})^2 + (d\Delta\alpha_{31})^2 + (d\Delta\alpha_{14})^2 + (d\Delta\alpha_{24})^2 + (d\Delta\alpha_{34})^2] \\ &\quad + 2f_{aa}[(d\Delta\alpha_{12})(d\Delta\alpha_{31} + d\Delta\alpha_{14} + d\Delta\alpha_{23} + d\Delta\alpha_{24}) + (d\Delta\alpha_{23})(d\Delta\alpha_{24} + d\Delta\alpha_{34} + d\Delta\alpha_{31}) \\ &\quad + (d\Delta\alpha_{31})(d\Delta\alpha_{14} + d\Delta\alpha_{34}) + (d\Delta\alpha_{14})(d\Delta\alpha_{24} + d\Delta\alpha_{34}) + (d\Delta\alpha_{24})(d\Delta\alpha_{34})] \\ &\quad + 2f_{da}[\Delta d_1(d\Delta\alpha_{12} + d\Delta\alpha_{31} + d\Delta\alpha_{14}) + \Delta d_2(d\Delta\alpha_{12} + d\Delta\alpha_{23} + d\Delta\alpha_{24}) \\ &\quad + \Delta d_3(d\Delta\alpha_{31} + d\Delta\alpha_{23} + d\Delta\alpha_{34}) + \Delta d_4(d\Delta\alpha_{14} + d\Delta\alpha_{24} + d\Delta\alpha_{34})], \end{aligned} \quad (55)$$

where d is the equilibrium C–H bond distance.

By the same procedure used for CH_3Cl , the following \mathbf{F} matrices were obtained: for type A_1 ,

$$(F_{11}) = (f_d);$$

for type E ,

$$(F_{11}) = (d^2[f_a - 2f_{aa}]);$$

and for type T_2 ,

$$\begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} = \begin{pmatrix} d^2f_a & \sqrt{2}df_{da} \\ \sqrt{2}df_{da} & f_d \end{pmatrix}.$$

$$\begin{aligned} (\sigma_d)R_{1a} &= AR_{1a} + BR_{1b} + CR_{1c}, \\ (\sigma_d)R_{1b} &= A'R_{1a} + B'R_{1b} + C'R_{1c}, \\ (\sigma_d)R_{1c} &= A''R_{1a} + B''R_{1b} + C''R_{1c}. \end{aligned}$$

The transformation matrix is therefore of the form

$$\begin{bmatrix} A & B & C \\ A' & B' & C' \\ A'' & B'' & C'' \end{bmatrix}.$$

Hence the character for the covering operation σ_d would be $A + B' + C''$, and this must be the same as the character for σ_d given in the table⁸ for the point group \mathbf{T}_d . The determination of the transformation properties of the symmetry coordinates for the other covering operations of the group would be carried out in the same way.

Wilson² has given some rules that may be used in choosing the coefficients for the symmetry coordinates in complicated cases.

The \mathbf{F} Matrices

For the present purpose it is desired to calculate force constants from the observed frequencies for the CH_4 molecule and subsequently to use these force constants to calculate the frequencies of the CD_4 molecule. Since there are only four distinct frequencies for CH_4 , only four force constants can be used in expressing the potential energy. The expression used was

In obtaining the \mathbf{F} matrix for the E vibrations, one may use either R_{1a} or R_{1b} ; and in obtaining the \mathbf{F} matrix for the T_2 vibrations, one may use any one of the pairs: R_{1a} , R_{2a} ; R_{1b} , R_{2b} ; R_{1c} , R_{2c} .

The \mathbf{G} Matrices

The elements for the \mathbf{G} matrix for the non-degenerate type A_1 vibration can be obtained from Eq. (20), and the \mathbf{G} matrix elements for the type E vibration can be obtained from Eq. (40). The corresponding equation for the ele-

TABLE VII. Fundamental frequencies for CH₄.

Type	Frequency (cm ⁻¹)	Designation
A ₁	3029.8	ν_1
E	1390.2	ν_2
T ₂	1357.6	ν_3
	3156.9	ν_4

ments of the **G** matrix for the type T₂ vibrations is²

$$G_{jl} = (1/3) \sum_p \mu_p g_p (\mathbf{S}_{ja}^{(t)} \cdot \mathbf{S}_{la}^{(t)} + \mathbf{S}_{jb}^{(t)} \cdot \mathbf{S}_{lb}^{(t)} + \mathbf{S}_{jc}^{(t)} \cdot \mathbf{S}_{lc}^{(t)}). \quad (56)$$

By proceeding in the same manner as for CH₃Cl, the following **G** matrices were obtained: for type A₁, (G₁₁) = (μ_H); for type E, (G₁₁) = (3 μ_H/d^2); and for type T₂,

$$\begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} = \begin{pmatrix} [2\mu_H + (16/3)\mu_C]/d^2 & -8\mu_C/(3d) \\ -8\mu_C/(3d) & \mu_H + 4\mu_C/3 \end{pmatrix}.$$

Force Constants for CH₄

The fundamental frequencies⁸ for CH₄ are given in Table VII. For the type A₁ vibration,

$$\lambda_1 = F_{11} G_{11} = (f_d)(\mu_H) = 4\pi^2 c^2 \nu_1^2, \quad (57)$$

whence

$$f_d = (4\pi^2 \nu_1^2)/\mu_H = 5.4514 \times 10^5 \text{ dyne/cm.}$$

For the type E vibration,

$$\lambda_2 = F_{11} G_{11} = [d^2(f_a - 2f_{aa})][3\mu_H/d^2] = 4\pi^2 c^2 \nu_2^2, \quad (58)$$

whence

$$f_a - 2f_{aa} = 0.38255 \times 10^5 \text{ dyne/cm.} \quad (59)$$

For the type T₂ vibrations, the secular equation in expanded form is

$$\lambda^2 - (F_{11}G_{11} + 2F_{12}G_{12} + F_{22}G_{22})\lambda + \begin{vmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{vmatrix} \cdot \begin{vmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{vmatrix} = 0. \quad (60)$$

Using

$$\mu_H = 5.9742 \times 10^{23} \text{ g}^{-1}, \quad \mu_C = 5.0183 \times 10^{22} \text{ g}^{-1}$$

and $d = 1.093 \times 10^{-8}$ cm, G₁₁, G₁₂, and G₂₂ were found to be 1.2241×10^{40} gm⁻¹ cm⁻², -1.2243×10^{31} gm⁻¹ cm⁻¹ and 6.6433×10^{23} gm⁻¹, respectively, and F₂₂ has already been obtained from Eq. (57). Substituting these values into Eq.

(60), one obtains

$$\begin{aligned} \lambda^2 - [(1.2241 \times 10^{40})F_{11} - (2.4486 \times 10^{31})F_{12} \\ + 3.6215 \times 10^{29}] \lambda + (4.3514 \times 10^{69})F_{11} \\ - (7.9821 \times 10^{63})F_{12}^2 = 0. \quad (61) \end{aligned}$$

From Eq. (25) and the frequencies given in Table VII,

$$\lambda_3 = 4\pi^2 c^2 (1357.6)^2 = 6.5389 \times 10^{28} \text{ sec}^{-2} \quad (62)$$

and

$$\lambda_4 = 4\pi^2 c^2 (3156.9)^2 = 3.5357 \times 10^{29} \text{ sec}^{-2}. \quad (63)$$

But, since λ_3 and λ_4 are roots of Eq. (61), one can write

$$(\lambda - \lambda_3)(\lambda - \lambda_4) = 0. \quad (64)$$

Expanding and substituting the values for λ_3 and λ_4 from Eqs. (62) and (63), one gets

$$\lambda^2 - (4.1896 \times 10^{29})\lambda + 2.3120 \times 10^{58} = 0. \quad (65)$$

Equating coefficients in Eqs. (61) and (65), one obtains the two relations

$$\begin{aligned} (4.3514 \times 10^{69})F_{11} - (7.9821 \times 10^{63})F_{12}^2 \\ = 2.3120 \times 10^{58}, \quad (66) \end{aligned}$$

$$(1.2241 \times 10^{40})F_{11} - (2.4486 \times 10^{31})F_{12} \\ = 5.6810 \times 10^{28}. \quad (67)$$

Eliminating F_{11} from Eqs. (66) and (67), one gets

$$F_{12}^2 - (1.0905 \times 10^{-8})F_{12} + 3.6652 \times 10^{-7} = 0. \quad (68)$$

Unfortunately, the values of F_{12} obtained from Eq. (68) are not real. This may be due to the fact that it was not possible to use the most general potential energy expression.

In order to obtain the values of the remaining force constants, it is therefore necessary to transfer f_a from the CH₃Cl molecule. Such a transfer of force constants between molecules of similar structure has been shown by Crawford and Brinkley⁹ to be possible for a number of molecules. Substituting this value, $f_a = 0.46 \times 10^5$ dyne/cm, into Eq. (67), one gets

$$F_{12} = \sqrt{2} f_{da} = 4.2702 \times 10^{-4} \text{ dyne}$$

or $f_{da} = 0.2763 \times 10^5$ dynes/cm. These values of f_a and f_{da} may be tested by substitution in Eq. (66).

Finally, the force constant f_{aa} can be obtained from Eq. (59).

Calculation of the CD_4 Frequencies

Substituting μ_D for μ_H in the \mathbf{G} matrices for the CH_4 molecule, and using the new matrix elements in Eqs. (57), (58) and (60), the frequencies for CD_4 were calculated. In Table VIII, the values obtained are compared with those calculated by Dennison⁸ using a different method. Despite the approximate potential energy expression used, the agreement is reasonably satisfactory. A somewhat better agreement might be obtained by slight adjustments in the force constants.

SUMMARY OF THE METHOD

It may be helpful to summarize briefly the principal steps in applying the Wilson method to the calculation of the vibrational frequencies of a molecule. First, orthonormal—that is, both orthogonal and normalized—symmetry coordinates, which are linear combinations of equivalent internal coordinates and which transform according to the characters of the vibration type concerned are formed. From the \mathbf{U} matrices obtained from these symmetry coordinates and from the \mathbf{f} matrix obtained from the potential energy expressed in terms of the internal coor-

TABLE VIII. Comparison of the present calculated frequencies of the CD_4 molecule with the values obtained by Dennison⁸ using a different method.

Type	Symbol	Vibration	Frequency (cm^{-1})	Dennison's value
		Present value		
A_1	ν_1	2151	2143.2	
E	ν_2	987	983.4	
T_2	ν_3	1019	1026.4	
	ν_4	2335	2336.9	

dinates, one gets the \mathbf{F} matrices for the different vibration types. In order to get the \mathbf{G} matrices, one first calculates the \mathbf{s}_{kt} vectors from the general expressions involving unit vectors directed along the chemical bonds. Next the $\mathbf{S}_j^{(t)}$ vectors are formed from the U_{jk} coefficients and the \mathbf{s}_{kt} vectors. The \mathbf{G} matrix elements are then obtained from the dot products of the $\mathbf{S}_j^{(t)}$ vectors. Finally, using these \mathbf{F} and \mathbf{G} matrices, the secular equations for the various vibration types are written in the expanded form, and the vibrational frequencies are then calculated from these equations.

* * *

The authors wish to express their appreciation to Mr. Leonard Reiffel for preparing the drawings for Figs. 1 to 5.

A BITTER paradox underlies many aspects of our era. The marvelous achievements of science are being used to destroy the civilization which science has achieved. Learning has been directed to the obliteration of its own temples. The great ideals of human liberty are led to the sacrifice by those who would preserve them. But of all the paradoxes around us none is more disconcerting than the fact that we, living during one of the greatest upheavals of recorded history, have so little understanding of the historical significance of our own times.—FELIX MORLEY.

Law of Elasticity for an Ideal Elastomer

ALLEN L. KING

Dartmouth College, Hanover, New Hampshire

ELASTOMERS are substances such as silicons, rubbers, plastics and other high polymers, having long chains of molecules. Many physiological properties of living tissues are determined by combinations of naturally occurring elastomers. In a recent article, Mark¹ has given an excellent survey of the chemical structure and physical texture of these materials.

Between the brittle and flow temperatures elastomers display a high degree of resilience, such as is commonly associated with natural rubber at room temperature. Since the stress-strain and thermoelastic characteristics are similar to those of rubber, these substances in this temperature range are said to be rubbery. The classical theory of elasticity based on Hooke's law is inappropriate for describing the elastic properties of these rubbery substances. That their elasticity is somehow related to the presence of long molecular chains was suspected at an early date, but this association did not become clearly formulated in a quantitative fashion until very recently.^{2,3} In the present paper a simplified and more consistent version of James and Guth's treatment is presented.

A block of unstretched elastomer is conceived to contain a large number of long molecular chains, so randomly twisted and intermeshed that any disturbance of them in one direction is readily communicated in all directions. For analytic purposes, this complex mesh is replaced by an equivalent network of chains extending between each of the three pairs of parallel surfaces. The number of molecules per unit length of chain and the number of chain-ends per unit area of surface are assumed to be constant throughout the block. Although chains having their ends in opposite surfaces are uniform

in length, the paths along the chains are very tortuous and variable. Each of these chains, if isolated, is supposed to coil up into a most probable configuration, owing to the thermal motions of molecules within the chain (internal Brownian movement). Thus a surface of the block may be considered pulled inward by the chains ending in it. So that there may be equilibrium in the unstretched state, this surface must be pushed outward by a force of equal magnitude. Such a force may be thought to arise from the sidewise thermal motions of chains that are more or less parallel to the surface.

If an external force X is applied so as to stretch the block in the x -direction, then the yz -surface is in equilibrium under the action of X , sidewise chain pressure and chain contraction. The resultant elongation of the individual chains ending in the yz -surface is believed to involve a reorientation of the intermolecular links, rather than any important change in intermolecular distances. If no volume change occurs, the internal energy remains constant under stress, and the applied force X equals $-T(\partial S/\partial x)$, where T is the absolute temperature and S is the entropy of the sample.^{2,4} Holt and McPherson⁵ have observed that the volumes of samples of several kinds of rubber remain practically constant up to as much as 300 percent elongation. If the volume of an ideal elastomer is assumed to be unaltered by stretching, then, for the aforementioned block, $V = x_0 y_0 z_0 = xyz$, where x_0, y_0, z_0 and x, y, z are the dimensions of the unstretched and stretched block, respectively. If the xz - and xy -surfaces are not constrained when force X is applied, then $y = y_0(x_0/x)^{\frac{1}{2}}$ and $z = z_0(x_0/x)^{\frac{1}{2}}$.

Consider a single chain ending in the yz -surface. It consists of a very large number of links N_z , each of length l . It pulls the surfaces together with force F , so that, if a small virtual elongation Δx were made, the potential energy gained by

¹ H. Mark, *Am. J. Physics* 13, 207 (1945).

² F. T. Wall, *J. Chem. Physics* 10, 485 (1942); 11, 527 (1943).

³ H. M. James and E. Guth, *J. Chem. Physics* 11, 455 (1943). See also E. Guth, "The problem of the elasticity of rubber and rubber-like materials" in *Surface chemistry*, F. R. Moulton, ed. (American Association for the Advancement of Science, Washington, 1943), pp. 103-127.

⁴ D. R. Elliot and S. A. Lippman, *J. App. Physics* 16, 50 (1945).

⁵ W. L. Holt and A. T. McPherson, *J. Research Natl. Bur. Standards* 17, 657 (1936).

the chain would equal $-F\Delta x$. Let $C(x)$ represent the number of chain configurations associated with end-separation x . Then the probability of finding the chain-ends separated by a distance between x and $x+dx$ is

$$pdx = \exp(F\Delta x/kT)C(x)dx, \quad (1)$$

where x equals $x_0 + \Delta x$. The probability function p has a maximum value when Δx is equal to the most likely extension e . An explicit relation between F and x may now be found, without determining the function $C(x)$, by evaluating $\exp(Fe/kT)$ for the chain and inserting the resulting expression in the equation

$$x = x_0 + kT\partial(\ln p_{\max})/\partial F. \quad (2)$$

In order to evaluate $\exp(Fe/kT)$, the link length l is considered to be a vector which may assume all possible orientations in space with equal probability. Because the number of links per chain is very large, the average contribution of each link to the exponential may be found and used. Let the i th link with component x_i along the x -axis make a contribution e_i to extension e of a typical chain ending in the yz -surfaces. Then

$$\begin{aligned} \exp(Fe/kT) &= \exp\left(\sum_{i, N_z} Fe_i/kT\right) \\ &= \prod_{i, N_z} \exp(Fe_i/kT). \end{aligned} \quad (3)$$

Extension e_i may assume values between $-l-x_i$ and $l-x_i$, so that, on the average,

$$\begin{aligned} \exp(Fe_i/kT) &= (1/2l) \int_{-l-x_i}^{l-x_i} \exp(Fe_i/kT) de_i \\ &= (\sinh u/u) \exp(-Fx_i/kT), \end{aligned} \quad (4)$$

where $u = Fl/kT$ and, therefore,

$$\exp(Fe/kT) = (\sinh u/u)^{N_z} \exp(-Fx_0/kT), \quad (5)$$

since $x_0 = x_1 + x_2 + \dots + x_i + \dots$.

On inserting this expression for $\exp(Fe/kT)$ in Eq. (2) and solving for x , one obtains

$$x = N_z l (\coth u - 1/u) = N_z l \mathcal{L}(Fl/kT). \quad (6)$$

The expression $\mathcal{L}(u)$ [$= \coth u - 1/u$], called the Langevin function, appears prominently in classical and modern theories of ferromagnetism.

Let q_z [$= 1/N_z l$] be a parameter characterizing the molecular chains with end-separation x . Then the contractile force per chain on the yz -surface is

$$F = (kT/l) \mathcal{L}^{-1}(q_z x), \quad (7)$$

wherein $\mathcal{L}^{-1}(q_z x)$ is defined as the inverse Langevin function of argument $q_z x$. If there are M chain-ends per unit area of surface in the unstretched block, the total inward pull on the yz -surface is $y_0 z_0 M F$. Similar equations for the contractile force of chains ending in the xy - and xz -surfaces may be set down.

The sidewise chain pressure P on the yz -surface may be thought to arise from thermal motions of molecules largely in chains extending between the xy - and xz -surfaces. This pressure, like that of a confined gas, is uniform over all surfaces of the block. Its value, therefore, may be obtained by examining the force situation at any one of the unconstrained surfaces. For instance, at the xy -surface the contractile force per unit area is $(MkT x_0 y_0 / lxy) \mathcal{L}^{-1}(q_z z)$, where q_z [$= 1/N_z l$] is the parameter for chains extending between the xy -surfaces. For equilibrium,

$$P = (MkT/l)(x_0/x)^{\frac{1}{2}} \mathcal{L}^{-1}(q_z z), \quad (8)$$

as will be seen upon recalling that $y = y_0(x_0/x)^{\frac{1}{2}}$ and $z = z_0(x_0/x)^{\frac{1}{2}}$. This same pressure is found on the xz -surface. Since the linear density of molecules and, therefore, of links has the same value for all chains and since the tortuosity of the chains in the unstretched block does not depend on the position of chain-ends, the number of links per chain is assumed to be proportional to the distance between ends, so that $q_z = q_y(y_0/z_0) = q_z(x_0/z_0)$.

If an external force X is applied to stretch the block along the x axis, evidently

$$\begin{aligned} X &= MFy_0 z_0 - Pyz \\ &= (MkT/l)y_0 z_0 [\mathcal{L}^{-1}(q_z x) \\ &\quad - (x_0/x)^{\frac{1}{2}} \mathcal{L}^{-1}\{q_z x_0(x_0/x)^{\frac{1}{2}}\}]. \end{aligned} \quad (9)$$

This is the desired relation between the applied force and the resulting extension along the line of action of the force. Division by $y_0 z_0$ yields the applied stress. If the arguments of the inverse Langevin functions are sufficiently small, the force per unit volume reduces to

$$X_v = (3Mq_z kT/l)[x/x_0 - (x_0/x)^2]. \quad (10)$$

Equation (10) with $x_0 = 1$ is to be compared with Wall's Eq. (31)² and with James and Guth's Eq. (6.12)³.

Even for small extensions of an elastomer the elastic force is not proportional to the elongation, as would be the case if Hooke's law applied. As a matter of fact, the general law, Eq. (9), yields a sigmoid curve, which is typical of rubber and other elastomers. Successful applications⁴ of this law to natural and synthetic rubbers lend support to the belief that the foregoing analysis is essentially correct. It should be noted, however, that for vulcanized and filled samples the effective chain lengths are reduced by introducing the concept of fix-points. These have the effect of decreasing the number of links per chain and, therefore, of increasing the q -value of the sample. Consequently, the sample becomes more rigid and less resilient.

From the general law for elastomers it is immediately evident that a larger force must be applied at elevated temperatures to produce the same amount of elongation as at room tempera-

ture. In other words, an elastomer behaves as though its effective Young's modulus were increased by a rise in temperature. This conclusion is in agreement with the well-known thermo-elastic properties of rubber. For quantitative applications, ordinary thermal expansion must be taken into account.⁶

The elasticity of animal tissues has long been recognized to be much like that of rubber.⁷ An extension of the foregoing analysis to cylindrical tubes with elastomeric walls yields a pressure-volume relation that describes fairly well experimental results on the human aorta.⁸ Furthermore, systems manifesting superposed elastic and viscous behavior have been studied in terms of the present theory.⁹ The general law for an ideal elastomer has been employed in a variety of problems with marked success.

⁶ S. L. Dart and E. Guth, *J. Chem. Physics* 13, 28 (1945).

⁷ C. S. Roy, *Foster's J. Physiol.* 3, 125 (1880).

⁸ A. L. King, to be published in *J. App. Physics* 17 (1946).

⁹ A. V. Tobolsky and R. D. Andrews, *J. Chem. Physics* 13, 3 (1945).

A Brief Account of the Physics of the Atomic Bomb

R. J. STEPHENSON
The College of Wooster, Wooster, Ohio

THE atomic bomb comes as a climax to an era in physics which may be said to have started about 50 years ago with the discovery of radioactivity by Becquerel. Underlying the theory of the release of atomic energy is the principle of the equivalence of mass and energy, $E = mc^2$, discovered by Einstein in 1905. During the early 1930's the rate at which fundamental discoveries were made in physics was considerably accelerated. Thus in 1932 Cockcroft and Walton effected the first artificial transmutation; Chadwick discovered the neutron, and Anderson the positron; Urey, Brickwedde and Murphy discovered deuterium; and the cyclotron and Van de Graaff machines were first developed. Artificial radioactivity was discovered by Curie and Joliot in 1933.

It was in 1934 that Fermi and his collaborators bombarded uranium with neutrons in the hope of

producing the transuranic elements of atomic numbers 93 and higher. Prior to this, these investigators had shown that nearly all elements could be made radioactive by bombardment with neutrons. The problem of identifying the products resulting from the bombardment of uranium by neutrons was by no means an easy one, for not only is the chemistry difficult but the amounts of the substances produced are extremely small. An account of the experiments of Fermi and others is contained in an article by Turner¹ on "Nuclear Fission."

The problem of producing transuranic elements was of course very exciting, and many physicists and chemists turned their attention to it as will be seen by looking at the long list of references given in the article by Turner. A

¹ L. A. Turner, *Rev. Mod. Physics* 12, 1 (1940).

quotation from this article shows how the apparently contradictory results of different workers were finally explained:

In a second paper [Hahn and Strassmann, January 1939] they gave the details of a most beautiful and thorough set of experiments which showed beyond a doubt that both "Ra III" and "Ra IV" were actually isotopes of barium rather than of radium. Further, the 15-min and the ~4 hr "Ra's" obtained from bombarded thorium were likewise shown to be isotopes of barium.

This was a type of disintegration previously unknown in nuclear physics; for now, in place of relatively light particles being ejected from an activated nucleus, the activated uranium and thorium nuclei had broken apart into comparatively heavy particles, one of which was identified as barium. The excitement which this discovery produced may be judged from the fact that nearly 100 papers on the subject were published in 1939.

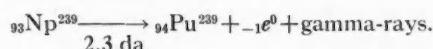
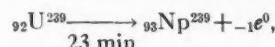
As is well known, the name *fission* was given to this phenomenon of the breakup of the uranium or thorium nucleus. Later it was shown that protoactinium could also show fission. The work in 1939 demonstrated that there was an enormous release of energy at fission—about 200 Mev per atom of uranium; that there were about 30 fission products associated with fission, barium being one of them; that these fission products were radioactive; that neutrons were also ejected on fission. Bohr and Wheeler² showed theoretically that the two important isotopes $_{92}\text{U}^{238}$ and $_{92}\text{U}^{235}$ behaved very differently towards fission and that it was the 235 isotope, having a relative abundance of 1 part in 140, which was primarily responsible for the fission phenomenon.

All this information appeared to some physicists to be of considerable importance to a world at war. The neutrons which are ejected at fission might perhaps be used to induce further fission and this in turn additional neutrons, thus producing a self-sustaining atomic chain reaction with an enormous release of energy. In other words, an atomic bomb might be possible. The British physicists, too, felt that a military weapon of vital importance might be based on the phenomenon of fission. Even with the war pressing them as it was in 1940, research in this

field was carried on in the British Isles. Fears were heightened when it was discovered that Germany was pursuing research in nuclear fission.

After it was decided to investigate further the possibility of the atomic bomb, nuclear physics became a secret field of investigation. As is well known, two official reports have recently been published on the work accomplished during the war.^{3,4} These documents represent the official release of information at present and, though a great amount of information is still secret, the interested reader will find it profitable to read them.

Bohr and Wheeler² predicted theoretically, and it was later confirmed experimentally, that the more common isotope of uranium, $_{92}\text{U}^{238}$, undergoes fission only when bombarded with high energy neutrons, and this to a relatively small extent; on the other hand, that the rarer isotope, $_{92}\text{U}^{235}$, would not only undergo fission to a small extent with high energy neutrons but also to a considerable extent with low energy, or thermal, neutrons. These qualitative statements concerning the relative extent of fission are expressed quantitatively in terms of cross section for fission, but their actual values still remain in the realm of secrecy. Another point of difference in the two isotopes is that $_{92}\text{U}^{238}$ can absorb neutrons of a "resonance" energy without undergoing fission. This resonance energy lies in the region above the thermal energies: from 1/40 to about 1000 ev. The probability of this kind of non-fission absorption in $_{92}\text{U}^{238}$ is very high. As was found by the Radiation Laboratory at the University of California, this nonfission absorption gives rise to the transuranic elements neptunium, atomic number 93, and plutonium, atomic number 94, through two successive beta-particle emissions as shown by the equations:



³ H. D. Smyth, *A general account of the development of methods of using atomic energy for military purposes under the auspices of the United States Government, 1940-1945* (Princeton Univ. Press); also, *Rev. Mod. Physics* 17, 351 (1945).

⁴ *Britain and the atomic bomb* (British Information Service, 1336 New York Ave., NW, Washington 5, D. C.); also, *Rev. Mod. Physics* 17, 472 (1945).

² N. Bohr and J. A. Wheeler, *Physical Rev.* 56, 426 (1939).

The 23 min and 2.3 da are the half-life times of the radioactive elements. Plutonium, $_{94}\text{Pu}^{239}$, does emit alpha-particles and so decays to $_{92}\text{U}^{235}$, but the rate of emission is so slow that plutonium is in effect a stable element. As regards fission, plutonium is very similar to the rarer isotope of uranium, $_{92}\text{U}^{235}$.

With this information available the policy committee in charge of the atomic-bomb project had to investigate the various means of obtaining rapidly either $_{92}\text{U}^{235}$ or $_{94}\text{Pu}^{239}$, or both, in quantities sufficient to make bombs. A great deal of attention was paid to the possible methods of separating $_{92}\text{U}^{235}$ from the natural material. This was by no means an easy task; a brief discussion of it will be given later.

First, a few of the problems involved in producing plutonium must be considered. This task was undertaken under the direction of A. H. Compton, as Director of the Metallurgical Laboratory of the University of Chicago, with E. Fermi, of Columbia University, and E. Wigner, of Princeton University, in charge of the experimental and theoretical work, respectively. To produce plutonium, neutrons must be absorbed by $_{92}\text{U}^{238}$. As was previously stated, this absorption takes place at neutron energies between approximately 1/40 and 1000 ev. The neutrons available for this resonance absorption must come from those released at fission, which come off with energies of the order of 1 Mev. Also, if this reaction is to be self-sustaining, at least one of the neutrons released from fission of each uranium atom must, after being slowed down to thermal energies, remain to produce fission in another uranium atom. Thus, if the reaction is to be self-sustaining, the reproduction factor, denoted by k , must be unity or greater.

As has just been mentioned, there must be some means for slowing down the fission neutrons to thermal energies. This slowing down may be accomplished by allowing the neutrons to collide with light atoms—the lighter the better, since the neutron mass is almost the same as that of the hydrogen atom. Thus, the elements at the beginning of the periodic table are the most effective in slowing down—in acting as “moderators”—for the neutrons. To get the reproduction factor k up to unity, so that a chain reaction may take place, these moderators must not absorb

neutrons. This requirement rules out hydrogen, lithium, and boron as moderators. Deuterium and beryllium were ruled out because of their unavailability; helium was ruled out because it is a gas and forms no compounds. Thus the most suitable moderator material was carbon, in the form of graphite blocks. This graphite had to be considerably purer than the usual commercial types.

The chain-reacting system used consisted of blocks of uranium or uranium oxide placed at some distance from one another in a lattice arrangement among the blocks of graphite. This was referred to as “a graphite pile,” or simply as a “pile.” In determining what geometric arrangement of uranium and graphite would produce the best value for k (greater than unity), it was not necessary to attempt to build a chain-reacting pile; instead, experiments could be made on a smaller unit called an “exponential” pile. In this unit a neutron source, radium-beryllium, was placed at the base of the uranium-graphite pile, and the thermal neutron intensity at various points in the pile was determined by measuring the activity of foils of some suitable material such as indium placed at these points. Since the neutron intensity decreases exponentially with increasing vertical distance from the neutron source, the piles were called exponential piles.

With the materials available, and with a particular geometric arrangement, Fermi reported a value of 0.87 for k in the fall of 1941. A lattice structure of graphite and uranium in the form of a cube about 8 ft on a side was used. The importance of a lattice structure lies in the fact that it reduces nonfission capture in $_{92}\text{U}^{238}$. Unless this is done the value of k would be too small for a chain reaction to take place. By July 1942 the value of the reproduction factor as calculated from the exponential pile experiments had exceeded unity; actually it was 1.007.

Before proceeding with a discussion of the chain-reacting pile, we shall briefly discuss the energy release and also the theory of a pile. The energy released on fission may be accounted for by the difference in the masses of the activated uranium nucleus and of the fission products into which the uranium nucleus splits. An approximate value of the energy given out may be obtained from the mass defects of the atoms in-

volved; it is calculated from the packing fraction curve. The packing fraction f of an atom is defined by the equation

$$f = (M - A)/A, \quad (1)$$

where M is the isotopic mass of the atom and A is the mass number, that is, the total number of protons and neutrons in the nucleus of the atom. If M_0 is the mass of the nucleus before fission, and M_1, M_2 are the nuclear masses of the fission products, then by Einstein's mass-energy relationship the energy ΔE released on fission is

$$\Delta E = c^2(M_0 - M_1 - M_2).$$

Introducing packing fractions from Eq. (1), we have

$$\Delta E = K[A_0(1+f_0) - A_1(1+f_1) - A_2(1+f_2)],$$

where f_0, f_1, f_2 are the packing fractions of the atoms of mass numbers M_0, M_1, M_2 , respectively, and K is a conversion factor which, if ΔE is to be in terms of 1 Mev, is approximately equal to 931. Since A_0 , the mass number of the activated uranium nucleus, must be equal to the sum of A_1 and A_2 , the mass numbers of the fission products, it follows that

$$\Delta E = K(A_0f_0 - A_1f_1 - A_2f_2).$$

Though the fission of every uranium atom does not give rise to the same two kinds of atoms, the fission products lie in two groups, a heavy one of atomic number ranging from 127 to 154, and a light one whose atomic number lies between 83 and 115; nevertheless, for these atoms the packing fractions are negative and approximately the same. Using as an example, $A_1=140$ and $A_2=96$ and using the packing fraction for these atoms and for uranium such as may be found approximately in the form of a curve in many textbooks,⁵ one finds that the release of energy for each fissionable atom is approximately 200 Mev.

For the heavy elements such as uranium, the mass number A is much larger than twice the atomic number Z ; that is, there are many more neutrons than protons in uranium. On the other hand, for stable atoms near the middle of the

periodic table, A is approximately equal to $2Z$. Thus, when the uranium nucleus divides, it cannot break down into two stable atoms. For instance, if one of the fission products is $^{56}\text{Ba}^{140}$, which is a stable isotope of barium, the other would have to be of atomic number 36 and atomic mass 96. The heaviest isotope of krypton, $Z=36$, has a mass of 86. Hence to attain stability there must be ejected from this latter nucleus a total of ten neutrons and electrons since the ejection of an electron is equivalent to the transformation of a neutron into a proton. The fission process then is accompanied by the emission of neutrons as well as by the production of highly beta-radioactive substances.

Now let us consider the theory of a chain-reacting pile by considering the life cycle of neutrons in a pile. Suppose at the start there are N high energy fission neutrons released in the graphite-uranium pile. On the average the energy of these N neutrons is above the threshold energy for producing fast fission in $^{92}\text{U}^{238}$ but, by collision with atoms of uranium and carbon, their energy is diminished and reaches the threshold value. Fission can be produced by some fast neutrons, and the original number N is increased by a small factor ϵ , which is called the *fast-fission effect*, or *fast multiplication factor*. The energy of the $N\epsilon$ neutrons is further reduced as the neutrons diffuse through the carbon and the resonance-energy capture region is reached. This neutron capture is taken into account by multiplying the number $N\epsilon$ of neutrons by a factor p , called the *resonance escape probability*, which is less than unity. The resonance escape probability is the probability that a given neutron starting with energy above the resonance region will reach thermal energy without capture in $^{92}\text{U}^{238}$. Thus of the original N high-energy neutrons there are $N\epsilon p$ neutrons that attain thermal energy.

Not all these neutrons are absorbed by the uranium since some are absorbed by the graphite and any other substance in the pile. There is a certain probability that a thermal neutron will be absorbed in the uranium, and this is expressed by a factor f called the *thermal utilization factor*. Thus there are $N\epsilon pf$ thermal neutrons absorbed in the uranium. Each thermal neutron absorbed by the uranium produces on the average η high energy fission neutrons. The final result is that

⁵ Richtmeyer and Kennard, *Introduction to modern physics* (McGraw-Hill), p. 590.

$N\epsilon\eta$ high energy neutrons are produced. If this number exceeds the original N , or if the quantity $\epsilon\eta$ is greater than unity, a self-sustaining chain reaction can take place. This assumes that there is no loss of neutrons by escape from the pile, which would be the case only if the pile were of infinite size. If k_∞ is the value of the reproduction factor for a pile of infinite size, then $k_\infty = \epsilon\eta$.

The critical size of a pile is that for which the production of free neutrons by fission is just equal to their loss by escape and nonfission capture. If the size is smaller than the critical size, no chain reaction can take place. In other words, k is less than unity for this pile.

We have already stated that in the exponential experiments a value for k greater than unity was obtained in the summer of 1942. By the fall of that year enough graphite, uranium oxide, and uranium metal were available to justify the construction of a self-sustaining pile. Since only 6 tons of pure uranium metal were available, the remaining lattice spaces had to be filled with the oxide. Instruments such as ionization chambers and Geiger counters filled with BF_3 or air were used to measure the neutron density and gamma-ray intensity as the pile was being constructed. Movable control rods made of cadmium metal and boron steel, both of which strongly absorb neutrons, were inserted in the pile so that the neutron density could be changed. Pushing in the control rods diminished the neutron density and lowered the value of the reproduction factor k . The pile began to operate on December 2, 1942 at a power of $\frac{1}{2}$ w. This was the first self-sustaining atomic chain reaction known to man! The power was then increased to 200 w but, since these experiments were being performed in a city, it was decided for reasons of safety not to increase the power further and the pile was dismantled. Early in 1943 a somewhat larger pile was built—the Argonne plant, just outside the city of Chicago. A 1000-kw air-cooled experimental pile was built at Clinton, Tennessee. These served as pilot plants for testing materials and processes for the production plant at Richland, Washington.

Once the uranium had been in a pile for a suitable time it had to be taken out and chemically treated for the extraction of the plutonium. In the Richland pile, which operated at a high

and secret power level, the uranium blocks were enclosed in tight aluminum cases and placed in long aluminum cylinders through which water flowed for cooling. Since the activated uranium contained all the fission products, each of which was highly radioactive, the chemical separations were by no means easy.

In the early days when plans were being made, it was decided to investigate as many means as possible for obtaining the rare isotope $_{92}\text{U}^{235}$ from natural uranium. Altogether, four methods for the separation of $_{92}\text{U}^{235}$ were considered, namely, gaseous diffusion, centrifugation, thermal diffusion, and the electromagnetic method. Only a pilot plant was built using centrifuges. Work on gaseous diffusion was begun at Columbia University in 1940 under H. C. Urey and J. R. Dunning. The principle of operation is that in a mixture of two gases of molecular weights M_1 and M_2 , respectively, the two components would diffuse through a porous barrier into an evacuated space at different rates.

Let us consider the gas, uranium hexafluoride, the two components of which, $_{92}\text{U}^{235}\text{F}_6$, $_{92}\text{U}^{238}\text{F}_6$, with molecular weights of 349 and 352, respectively, diffuse at different rates. In the ideal case the rates of diffusion are inversely proportional to the square roots of the molecular weights. If a small quantity of uranium hexafluoride diffuses through a porous barrier, the diffusate is enriched in $_{92}\text{U}^{235}\text{F}_6$ by a factor of $(352/349)^{\frac{1}{2}}$, or 1.0043. For an ordinary sample of uranium the relative concentrations of the 235 and 238 isotopes is 1 to 140. If a relative concentration of 9 to 1 is desired, there has to be an enrichment factor of 9×140 , or 1260. If the ideal enrichment factor on passing once through a single barrier is 1.0043, then it can be seen that there must be a considerable number of multiple-stage recycling diffusion units in order to reach the relative concentration of 9 to 1. A large production plant covering several acres and working on the gaseous diffusion principle was built at Clinton, Tennessee.

The other method of isotope separation which proved successful was the electromagnetic one for which E. O. Lawrence, of the University of California, was responsible. In this method the ions of the uranium isotopes are accelerated in an electric field, and the beam is bent by a magnetic

field so that the 235 isotope is brought to a focus at a different place from the 238 isotope, as in the electromagnetic spectrograph. If the limiting slits in the apparatus are narrow, the enrichment factor will be high but the yield will be small. A large number of production units were put into operation at Clinton, Tennessee. This was the first of the methods to produce the 235 isotope in quantity. By starting with uranium which was already enriched somewhat in the 235 isotope, the yield is considerably increased. As stated in the Smyth report, "an electromagnetic unit that could produce one gram a day of 40 percent pure U-235 from natural uranium could produce two grams a day of 80 percent U-235 if the concentration of U-235 in the feed material was twice the natural concentration (1.4 percent instead of 0.7 percent)." Owing largely to the work of P. H. Abelson, of the Naval Research Laboratory, a thermal diffusion⁶ plant was set up and operated for the purpose of supplying enriched feed material to the electromagnetic separator. This increased the production rate of $_{92}\text{U}^{235}$ considerably.

The purpose of all this work during wartime was to forge the powerful weapon which has come to be known as the atomic bomb. This work was placed in the charge of J. R. Oppenheimer, of the University of California, and was eventually located at Los Alamos near Santa Fe, New Mexico. What is needed in a bomb is a sudden release of energy. An atomic bomb has to be so constructed that a large fraction of the uranium nuclei undergo fission before the materials composing the bomb become appreciably separated. Separation of the uranium would mean that neutrons would escape, and thus the chain reaction would stop before all the fissionable material was consumed. Fission in this case must be caused by fast neutrons, for the time duration of the chain reaction, if it is to be efficient, must be extremely small. Just as in the huge graphite-uranium pile, where graphite was used on the outside as a neutron reflector to reduce the

critical size of the pile, so a reflector may be used to reduce the critical size of the bomb. In the latter case the reflector, known as a "tamper," is also useful in reducing the rate of expansion of the bomb by virtue of its inertia.

Unless the amount of $_{92}\text{U}^{235}$ or of Pu^{239} used in the bomb is larger than the critical amount, no chain reaction takes place; and when the amount is larger than the critical amount, the chain reaction spontaneously takes place, for there are always a few neutrons available to initiate the reaction. The detonation of the bomb would then appear to be effected by bringing subcritical amounts of the fissionable material into sudden and intimate contact. It should be mentioned that a chain reaction cannot be produced in a block of pure uranium metal, no matter how large, because the resonance or nonfission capture of neutrons in $_{92}\text{U}^{238}$ is relatively large. The actual construction of the bomb is still veiled in secrecy, as is the efficiency of the bomb.

An article of this kind can touch only on the high points of the atomic bomb. It has been suggested that 30 volumes be devoted to the report of the work of the Metallurgical Laboratory at the University of Chicago alone. This gives some idea of the encyclopedic nature of an account of the whole project.

By making and using an atomic bomb, physicists have created a new problem in the world. Mankind has to decide what to do with this enormously destructive weapon. Since nuclear energies are of the order of millions of electron volts while chemical energies are of the order of a few electron volts, it follows that this weapon is approximately a million times as potent as the usual type of explosive weapon. Mankind will have to make the choice of either world cooperation of an effective kind or of partial world annihilation. Without effective world cooperation scientists may find themselves involved in an armament race in nuclear physics. Scientists have created this weapon and have a moral obligation to see that their work is used for the betterment of civilization.

⁶ Brief descriptions of this process are given in reference 3 and in Pollard and Davidson, *Applied nuclear physics* (Wiley), p. 180.

Geometric Interpretation of the Character of Electron Diffraction Patterns

GENE F. ROSE

Department of Mathematics, University of Wisconsin, Madison Wisconsin*

THE fundamental phenomenon from which electron diffraction arises is, of course, the scattering of an electron beam by an elementary particle of matter. The atom is not the ultimate scattering particle, or center; the treatment of atomic scattering is inevitably concerned with the fine structure of the atom. Studies on this topic are numerous, and those cited¹⁻⁸ are merely representative.

The problem of electron diffraction can be approached quite rigorously by direct recourse to the theory of the fundamental phenomenon. A much simpler treatment, however, can be used with good approximation. For this purpose, the electron beam is replaced by a periodic space function with a wavelength λ equal to the de Broglie wavelength⁹ of the beam. In what follows, the beam is assumed to be monochromatic and parallel. Furthermore, it is assumed that an atom in the beam becomes a secondary point source, from which is propagated a spherical wave of unchanged wavelength. The phase of the disturbance is continuous at the secondary source.

Under these simplifying conditions, a geometric development of the laws of electron diffraction can be effected in the manner usually adopted in the study of the diffraction of light.¹⁰

* The author prepared this paper while he was associated with the Research Laboratories of the Westinghouse Electric Corporation.

¹ F. Kirchner, "Some new experiments on electron diffraction," *Physik. Zeits.* **31**, 1025-8 (1930).

² H. A. Bethe, et al., "Multiple scattering of electrons," *Proc. Am. Phil. Soc.* **78**, 573-85 (1938).

³ J. A. Wheeler, "The problem of the scattering of fast electrons," *Physical Rev.* **57**, 352-3 (1940).

⁴ M. Jauch, "The scattering problem in the electron pair theory," *Helv. Physica Acta* **13**, 451-72 (1940); **15**, 221-32 (1942).

⁵ L. A. Kul'chitskii, et al., "The scattering of fast electrons," *J. Physics (USSR)* **5**, 249-61 (1941).

⁶ H. E. Farnsworth, "Fine structure in the diffraction of low-speed electrons," *Physical Rev.* **60**, 168 (1941).

⁷ P. Urban, "The scattering of fast electrons by nuclei," *Zeits. f. Physik* **119**, 67-78 (1942).

⁸ A. I. Andrievskii, et al., "Multiple scattering of fast electrons," *J. Exp. Theoret. Physics (USSR)* **12**, 16-21 (1942).

⁹ L. de Broglie, *Thèse* (Mason & Cie., Paris, 1924); *Ann. physique* **3**, 22 (1925).

¹⁰ C. F. Meyer, *The diffraction of light, x-rays, and material particles* (Univ. of Chicago Press, 1934), pp. 29-40.

Line Grating Diffractions

A useful elementary grouping of diffracting atoms is that constituting a line grating and consisting of N atoms P_i ($i = 1, 2, \dots, N$), equivalent by translation \mathbf{a} (see Fig. 1). At any point in space, the diffraction is the resultant cooperative scattering from all N atoms. In so far as the scattered wavelets can be considered approximately spherical, the diffraction is symmetric with respect to the line of the grating. As a result, its determination in the plane containing beam direction and grating is completely general.

In Fig. 2, let \mathbf{a} be the unit translation and also the positive grating direction.** (Fig. 3.) If it is assumed that P_1Q is large compared to P_1P_r , the following approximate relations obtain among the contributions from the atoms of the grating. At any point Q in space, the contributions are (1) equal in amplitude, (2) collinear and (3) different in phase by $(r-1)\phi$ from that due to the first atom, where

$$\phi = (2\pi a/\lambda)(\cos D - \cos I), \quad (1)$$

where r is the order on the grating of the contributing atom, and D is the common limiting value of θ_1 and θ_2 .

Collinear wave motions compound vectorially.¹⁰ Thus if A_0 is the common amplitude of the waves at Q , the vector representing the resultant wave

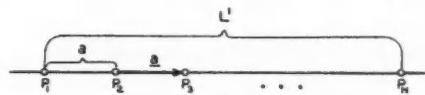


FIG. 1. The line grating.

** As indicated in Fig. 3, we shall arbitrarily choose the positive direction on the line grating and shall represent it by a vector \mathbf{G} . The direction of propagation of the incident beam is \mathbf{I} , and that of any diffracted ray is \mathbf{D} .

The magnitudes I and D of the rotations taking \mathbf{G} into \mathbf{I} and \mathbf{D} , respectively, are defined as the angles of incidence and diffraction, respectively. The magnitude δ of the rotation taking \mathbf{I} into \mathbf{D} is defined as the angle of deflection.

If \mathbf{I} , \mathbf{G} and \mathbf{D} are not coplanar, all these angles are signless, and they are to be evaluated in the interval $(0, \pi)$. If, however, these three directions are coplanar, δ will be taken as the algebraic difference $D - I$.

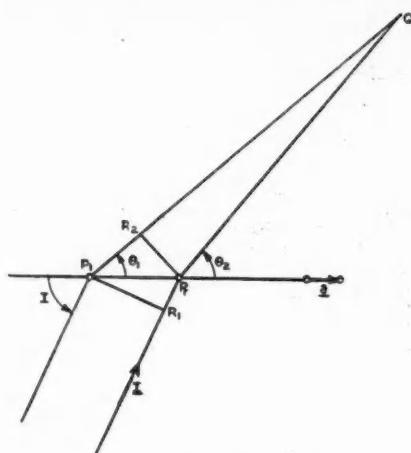


FIG. 2. Cooperatively scattering atoms.

is, in complex representation,

$$\sum_{r=1}^N A_0 e^{i(r-1)\varphi}. \quad (2)$$

Summing this geometric progression, we find that it can be written

$$A_0 \frac{\sin \frac{1}{2} N \varphi}{\sin \frac{1}{2} \varphi}, \quad (3)$$

which corresponds to a wave amplitude

$$A = A_0 \left| \frac{\sin \frac{1}{2} N \varphi}{\sin \frac{1}{2} \varphi} \right|, \quad (4)$$

and a phase

$$\Phi = \frac{1}{2}(N-1)\varphi \quad \text{or} \quad \Phi = \frac{1}{2}(N-1)\varphi + \pi, \quad (5)$$

according as the ratio of sines in expression (3) is positive or negative.

When $\sin \frac{1}{2} \varphi = 0$, A is NA_0 . By a familiar theorem concerning complex numbers,

$$A \leq NA_0, \quad (6)$$

so that the values

$$\phi = 2n\pi \quad (7)$$

(n any integer or zero) correspond to the greatest maximum of amplitude. From Eq. (1) it is obvious that these maximums are defined by a Laue equation,¹¹

$$\cos D - \cos I = n\lambda/a, \quad (8)$$

¹¹ Reference 10, pp. 322, 452.

where n is now defined to be the *order* of the corresponding diffracted ray.

There are *extinctions*—that is, the amplitude vanishes—when

$$\sin \frac{1}{2} N \varphi = 0,$$

$$\sin \frac{1}{2} \varphi \neq 0, \quad (9)$$

hence when

$$\phi = 2\pi p/N, \quad (10)$$

where p is an integer not divisible by N .

From Eqs. (7), (10) and (1), it is evident that the relation

$$\cos D - \cos I = \left(n + \frac{\theta}{N} \right) \frac{\lambda}{a}, \quad |\theta| < 1 \quad (11)$$

defines a *D*-interval for fixed *I*, or an *I*-interval for fixed *D*, in the neighborhood of the *n*th-order maximum, for which the scattering is not extinguished. We shall call this the *n*th-order interval of resolution of the variable concerned. This restriction of scattering to a region in space, due to periodicity in a direction \mathbf{a} , may be referred to as the \mathbf{a} -resolution.

Figure 4 shows the relative amplitude A/A_0 as a function of the phase change ϕ per unit spacing.

Figure 5 shows the behavior of scattering for incidence and diffraction directions nearly parallel

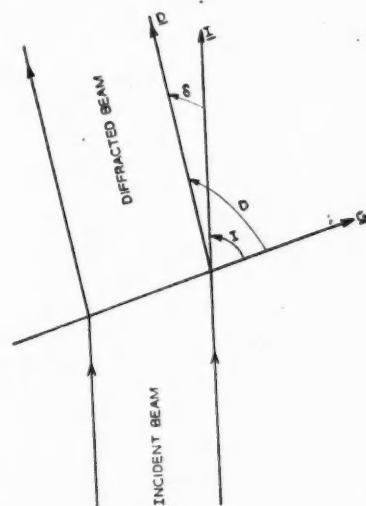


FIG. 3. Identification of angles and directions associated with line grating diffraction.

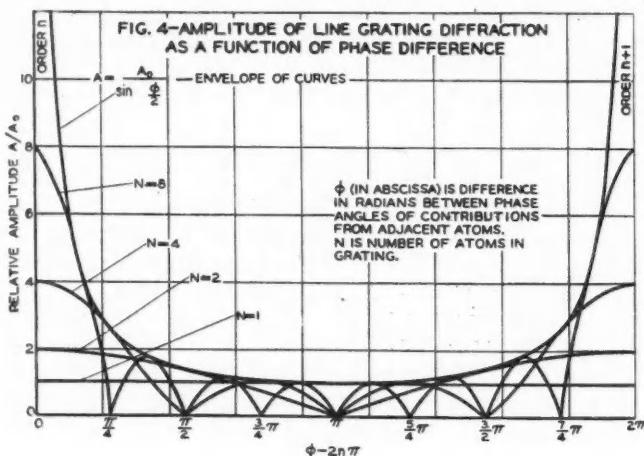


FIG. 4. Amplitude of line grating diffraction as a function of phase difference.

to the grating. Here the relative intensity E/E_0 , which is A^2/A_0^2 , or

$$E = E_0 \sin^2 \frac{1}{2}N\varphi / \sin^2 \frac{1}{2}\varphi, \quad (12)$$

is plotted as a function of D for several fixed values of I . Typical numerical values are used for the construction of this graph; namely, $\lambda = 0.05\text{A}$, $a = 10\text{A}$ and $N = 10$.

Each curve of the family in Fig. 5 has a peak at the Laue maximum and a first extinction on one or both sides of this value. Setting $n = 0$, $\theta = \pm 1$ in Eq. (11), we find for the values D_1 and D_2 of D at first extinction about the central (zero-order) maximum,

$$D_1 = (I^2 - 2\lambda/L)^{\frac{1}{2}}, \quad (13)$$

$$D_2 = (I^2 + 2\lambda/L)^{\frac{1}{2}}, \quad (14)$$

an approximation valid only for small angles. Here

$$L = Na = L' + a, \quad (15)$$

where L' is the length of the diffracting grating. At near-normal incidence, where both $D_1 - I$ and $D_2 - I$ are small in comparison with I , we must use a different approximation, namely,

$$D_1 = I - (\lambda/L \sin I), \quad (16)$$

$$D_2 = I + (\lambda/L \sin I). \quad (17)$$

Thus, from Eq. (13), if $I < (2\lambda/L)^{\frac{1}{2}}$, D_1 has no real value, and the zero-order diffraction fills a cone of half-angle D_2 about the grating axis; otherwise it is limited to the space between two such cones of half-angles D_1 and D_2 .

Bragg¹² gives $(2\lambda/L')^{\frac{1}{2}}$ for the angular radius of the central spot, and Thomson¹³ gives $\lambda/(L' \sin I)$ as the general angular half-width of a zero-order diffraction.

After the first extinction, any intensity curve exhibits a series of subsidiary maxima. These are displayed only for the parallel-incidence curve in Fig. 5. From Eq. (4), we see that these maxima satisfy the conditions

$$N \sin \frac{1}{2}\varphi \cos \frac{1}{2}N\varphi - \cos \frac{1}{2}\varphi \sin \frac{1}{2}N\varphi = 0, \quad (18)$$

$\sin \frac{1}{2}\varphi \neq 0$.

For $N > 2$, $\cos \frac{1}{2}N\varphi = 0$ satisfies Eq. (18) to a good approximation, whence, from Eqs. (4) and (10), we have for the p th subsidiary maximum ($p \leq N - 2$),

$$E_p = \frac{E_0}{\sin^2(2p+1)\pi/2N}. \quad (19)$$

The first maximum is evidently the greatest, so that it determines what we might call the upper level of *background scattering*. This level, the ratio of E_1 to the intensity $N^2 E_0$ of the integral-order maximum is, for large N , approximately $1/(9\pi^2/4)$, or less than 5 percent. This agrees with the observation of Meyer.¹⁴

¹² W. L. Bragg and F. Kirchner, "The action of a crystal as a two-dimensional lattice in diffracting electrons," *Nature* **127**, 738 (1931).

¹³ G. P. Thomson, "The small-scale structure of surfaces," *Phil. Mag.* **18**, 648 (1934).

¹⁴ Reference 10, p. 89 (note).

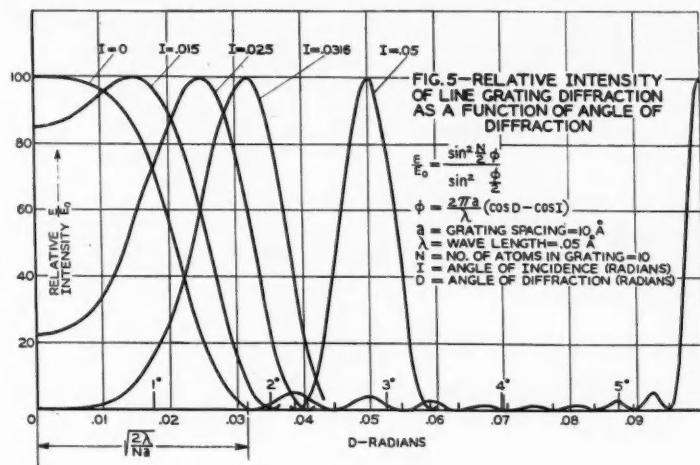


FIG. 5. Relative intensity of line grating diffraction as a function of angle of diffraction.

The Grouping of Line Gratings.

We have considered the line grating as a group of scattering points equivalent by a translation \mathbf{a}_1 . A plane grating can in turn be looked upon as a group of such line gratings equivalent by a translation \mathbf{a}_2 not collinear with \mathbf{a}_1 . Finally, a three-dimensional grating can be regarded as a group of plane gratings equivalent by a third, noncoplanar translation \mathbf{a}_3 .

By compounding, as before, the contributions of individual line gratings, and then of plane gratings, we can arrive successively at the diffracting properties of two- and three-dimensional lattice arrays of atoms. We now define I_i and D_i as the angles made by \mathbf{I} and \mathbf{D} , respectively, with the translation vector \mathbf{a}_i . Then N_i denotes the extent of the grating in direction \mathbf{a}_i , in number of atoms; L_i is $N_i a_i$. Then the intensity of scattering is

$$E = E_0 \prod_i \frac{\sin^2 \frac{1}{2} N_i \varphi_i}{\sin^2 \frac{1}{2} \phi_i}, \quad (20)$$

where

$$\varphi_i = \frac{2\pi a_i}{\lambda} (\cos D_i - \cos I_i). \quad (21)$$

The integral-order maximums are defined by the simultaneous Laue equations,

$$\cos D_i - \cos I_i = n_i \lambda / a_i. \quad (22)$$

The diffracted rays in the neighborhood of the maximum defined by the set of integers (n_i) are

restricted to a region in space defined by the simultaneous equations

$$\cos D_i - \cos I_i = \left(n_i + \frac{\theta_i}{N_i} \right) \frac{\lambda}{a_i}, \quad |\theta_i| < 1. \quad (23)$$

Equations (20) to (23) can be applied to a grating of v dimensions by letting i range from 1 through v .

Plane Grating Diffractions.

We might expect equations of the type (23) to be useful in the analysis of electron diffraction patterns from thin crystals. This problem has been of considerable interest. The work of Linnik,¹⁵ with x-rays, and of Kikuchi,¹⁶ with electrons, on thin sheets of mica raise a question as to whether thin crystals may not consist, under certain conditions, of an assemblage of two-dimensional gratings, with the periodicity in the third dimension somehow destroyed. Furthermore, the grazing-incidence, or "reflection," type of electron diffraction from unpolished surfaces is widely accepted¹⁷ to be a transmission phenomenon involving out-jutting thin crystals.

Actually, it is difficult to visualize the quantitative characteristics of thin-crystal diffraction with the aid of these equations. In the following

¹⁵ W. Linnik, "Diffraction of x-rays by two-dimensional crystal lattices," *Nature* 123, 604-5 (1929).

¹⁶ S. Kikuchi, *Japan J. Physics* 5, 83 (1928).

¹⁷ G. I. Finch, et al., "Electron diffraction and surface structure," *Trans. Faraday Soc.* 31, 1071 (1935).

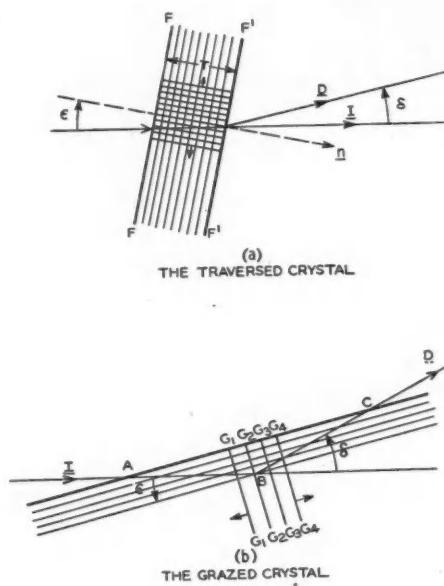


FIG. 6.

analysis, the introduction of the concept of null line gratings appears to be a simplification somewhat analogous to the idea of Bragg reflecting planes¹⁸ in x-ray diffraction analysis.

The thin crystal that is traversed by an electron beam—be it a large, single crystal, as of mica, or a microscopic one projecting from a “smooth” surface—might be regarded ideally as a set of plane gratings [in Fig. 6(a), through the fine lines parallel to FF' and $F'F'$ and normal to the plane of the paper]. The extent in these planes can usually be regarded as infinite, in so far as the resolution by spacings in them is concerned. Owing to the limited penetrating power of the electron beam, however, the extent in the direction of the beam of the set of planes is limited to the order of 100 Å, or perhaps 10 to 100 spacing units.¹⁹

To illustrate the method of attack, we shall consider a space lattice such that the plane gratings involved are equivalent by translation in the direction \mathbf{n} of their normal. The planes through FF' and $F'F'$ are boundary faces, and the

beam is assumed to enter and emerge through opposite faces.

Let us choose two spacing vectors (\mathbf{a} and \mathbf{b} , Fig. 7) in one of the plane gratings. In the figure they are referred to the orthogonal unit vectors \mathbf{u} , \mathbf{v} and \mathbf{n} , where \mathbf{v} is collinear with \mathbf{b} , \mathbf{n} is normal to the grating plane, and \mathbf{u} makes an acute angle θ with \mathbf{a} . The incident beam may be defined by the angles I_a and I_b made by the unit incidence direction vector \mathbf{i} with \mathbf{a} and \mathbf{b} , respectively; any diffracted beam, by the angles D_a and D_b made by the unit diffraction direction vector \mathbf{d} with the same respective spacing vectors. By our assumptions, these angles satisfy unmodified Laue conditions [Eqs. (22)], and we define a particular diffracted beam by the pair of component orders (n_a, n_b) .

From the appropriate Laue equations, we find, after substituting for I_a , I_b , D_a and D_b ,

$$\cos \theta (\cos \alpha_d - \cos \alpha_i) + \sin \theta (\cos \beta_d - \cos \beta_i) = n_a \lambda / a, \quad (24)$$

$$\cos \beta_d - \cos \beta_i = n_b \lambda / b. \quad (25)$$

Now let AB be any line making equal angles with \mathbf{i} and \mathbf{d} . The assumed equality of angles leads to the equivalent relations

$$\mathbf{t} \cdot (\mathbf{d} - \mathbf{i}) / t = 0, \quad (26)$$

$$(\cos \alpha_d - \cos \alpha_i) \cos \psi + (\cos \beta_d - \cos \beta_i) \sin \psi = 0, \quad (27)$$

where \mathbf{t} is any vector collinear with AB .

We are led by Eqs. (24), (25) and (27) to the conclusion,

$$n_a b / n_b a = \sin \theta - \tan \psi \cos \theta = OB/OA'; \quad (28)$$

in other words, that the lattice indices (h_0, k_0) of AB are in the ratio

$$h_0/k_0 = n_a/n_b. \quad (29)$$

The indices h_0 and k_0 refer to the \mathbf{a} and \mathbf{b} axes, respectively.

The lattice lines of the set (n_a, n_b) therefore fulfil the condition of making equal angles with the incident and diffracted beams. The contributions of the corresponding individual line gratings are thus zero-order diffractions, regardless of the incidence direction. Let us say that the set is *null* with respect to the diffraction (n_a, n_b) ; that its members are the *null line gratings corre-*

¹⁸ M. J. Buerger, *X-ray crystallography* (Wiley, 1942), p. 43.

¹⁹ Reference 13.

sponding to that diffraction in the particular set of planes.

The spacing of the set of null lines is readily calculated to be

$$s = -eb \cos \psi / n_b, \quad (30)$$

where e is the highest common factor of n_a and n_b .

Let P be an atom on a given null line, and P' any atom in the next null line out from the origin. If \mathbf{r} denotes the displacement PP' , we have from Eq. (1) that the contributions from the two atoms differ in phase by

$$\phi_0 = \frac{2\pi}{\lambda} \mathbf{r} \cdot (\mathbf{d} - \mathbf{i}). \quad (31)$$

But

$$\mathbf{r} = \mathbf{s} + \mathbf{t}, \quad (32)$$

where \mathbf{s} is the spacing vector of the null lines, and \mathbf{t} is the component of \mathbf{r} in their direction. Then, in view of Eq. (26),

$$\phi_0 = \frac{2\pi}{\lambda} \mathbf{s} \cdot (\mathbf{d} - \mathbf{i}). \quad (33)$$

Now \mathbf{s} has the direction cosines $(\sin \psi, -\cos \psi, 0)$ and length s . By making the appropriate substitutions in Eq. (33) from Eqs. (25), (27) and (30), we find

$$\phi_0 = \frac{-2\pi n_b s}{b \cos \psi} = 2\pi e. \quad (34)$$

It follows at once that the contributions of successive null line gratings differ in phase by $2\pi e$. Now we have the following useful conception of any plane-grating diffraction: *Any diffraction from a plane grating consists of the zero-order diffractions of the individual null line gratings, resolved in the e th order by their spacing.*

The deflection δ [Fig. 6(a)] of any diffracted beam measures some spacing of the crystal lattice. It varies, however, with the incidence direction. Let us now refer \mathbf{i} and \mathbf{d} to the orthogonal vectors \mathbf{s} , \mathbf{t} and \mathbf{n} (of Fig. 7), with respect to which they have the direction cosines (μ_a, μ_t, μ_n) and (ν_s, ν_t, ν_n) , respectively. Then

$$\cos \delta = \mu_a \nu_s + \mu_t \nu_t + \mu_n \nu_n. \quad (35)$$

But, from Eqs. (33) and (34),

$$\nu_s = \mu_s + (\lambda e / s); \quad (36)$$

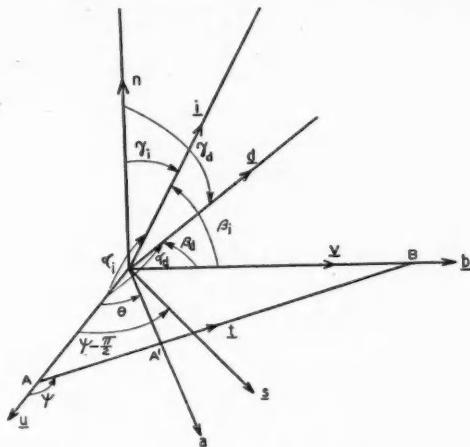


FIG. 7. Vectors associated with the null lines of a plane grating diffraction.

whereas, by hypothesis, $\nu_t = \mu_t$. Finally,

$$\begin{aligned} \nu_n &= \pm (1 - \nu_s^2 - \nu_t^2)^{\frac{1}{2}} \\ &= \pm \left(\mu_n^2 - 2 \frac{\lambda e}{s} \mu_s - \frac{\lambda^2 e^2}{s^2} \right)^{\frac{1}{2}}. \end{aligned} \quad (37)$$

Hence

$$\begin{aligned} \cos \delta &= 1 - \mu_n^2 + \frac{\lambda e}{s} \mu_s \\ &\quad \pm \mu_n \left(\mu_n^2 - \frac{2 \lambda e}{s} \mu_s - \frac{\lambda^2 e^2}{s^2} \right)^{\frac{1}{2}}. \end{aligned} \quad (38)$$

Let us consider a simple case of the variation of δ with incidence direction. Suppose that a plane grating is made up of a set of line gratings intersecting the plane of the paper normally in L_1, L_2, \dots (Fig. 8). The incidence direction \mathbf{I} will be taken in the plane of the paper. The given set of line gratings is null with respect to diffractions in the plane of the paper. We shall turn our attention to those diffractions that are of modular first order with respect to the spacing \mathbf{s} of the null line gratings; in particular, to those two of the "transmitted" or low-deflection type. One of these has the direction \mathbf{D}_{-1} . For its deflection δ_{-1} the general formula (38) is replaced by

$$\delta_{-1} = \epsilon + \sin^{-1} \left(\frac{\lambda}{s} \sin \epsilon \right), \quad (39)$$

which is obtained from Eq. (36) for $e = -1$. We agree to evaluate the indicated arc sine in the

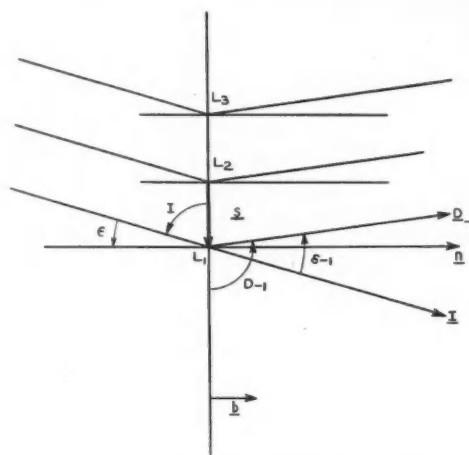


FIG. 8. Resolution of null line grating diffractions.

interval $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$. Likewise, for the other modular first-order diffraction of this type,

$$\delta_1 = \epsilon - \sin^{-1} \left(\frac{\lambda}{s} + \sin \epsilon \right), \quad (40)$$

with the same restriction on the arc sine function.

Figure 9 illustrates the manner of variation of these two diffractions as the incidence direction is rotated in the plane of the paper. Typical numerical data are adopted. The complete graph is symmetric with respect to the origin.

Analysis of Eqs. (39) and (40) shows that the minimum absolute deflection $|\delta^0|$ is the same for both modular first-order diffractions. It occurs when $\delta = 2\epsilon$, hence when the beam is apparently reflected from the set of planes containing the null line gratings and normal to their plane. For this condition, Eqs. (39) and (40) reduce to the Bragg relation²⁰

$$\lambda = 2s \sin \frac{1}{2} |\delta^0|. \quad (41)$$

For small values of ϵ , Eq. (39) or (40) gives the approximate relation

$$|\delta| = \lambda/s, \quad (42)$$

the same approximation which Eq. (41) yields for δ^0 .

As the incident beam approaches parallelism with the plane of the grating, $|\delta|$ can be shown

²⁰ W. L. Bragg, *Proc. Camb. Phil. Soc.* 17, 43 (1912); reference 10, pp. 292-4.

from Eq. (39) or (40) to approach the approximate value,

$$|\delta|_{\max} = (2\lambda/s)^{\frac{1}{2}} = (2|\delta^0|)^{\frac{1}{2}}. \quad (43)$$

We shall not here examine in detail the effect on δ of the unrestricted rotation of the incidence direction. However, we shall conclude that, in general, the deflection has values approximated by Eq. (42) for near-normal incidence, and that it becomes much greater for incidences which are less nearly normal.

Resolution of Plane Grating Contributions

Returning to our hypothetical thin crystal, let us suppose that it is built up of a set of N_b planes like the one indicated in Fig. 8, equivalent by the translation b in the plane of the paper. The effect of the b -resolution, as we have seen from our treatment of line grating diffractions, is to restrict the contributions from the individual plane gratings either to the interior of a cone or to the space between two cones, depending upon the incidence direction. There is such a region for each order of b -resolution. We shall deal only with that of the zero order.

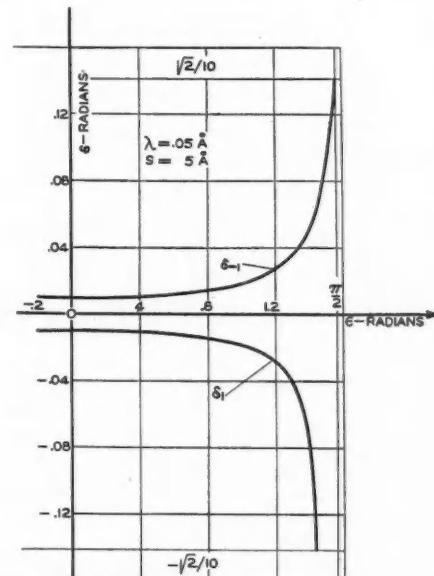


FIG. 9. Angle of deflection, δ , of plane grating diffraction as a function of angle of incidence, $\frac{1}{2}\pi - \epsilon$. See Eqs. (39) and (40).

The condition imposed by this added resolution on the deflection δ of a diffraction in the plane of the paper is, from Eq. (11),

$$|\cos(\delta - \epsilon) - \cos \epsilon| < \lambda/L_b. \quad (44)$$

The resulting δ -interval of resolution is composed, in general, of the sub-intervals (Δ_1, Δ_2) and (Δ_3, Δ_4) , where

$$\Delta_1 = \epsilon - \cos^{-1} \left(\cos \epsilon - \frac{\lambda}{L_b} \right), \quad (45)$$

$$\Delta_2 = \epsilon - \cos^{-1} \left(\cos \epsilon + \frac{\lambda}{L_b} \right), \quad (46)$$

$$\Delta_3 = \epsilon + \cos^{-1} \left(\cos \epsilon + \frac{\lambda}{L_b} \right), \quad (47)$$

$$\Delta_4 = \epsilon + \cos^{-1} \left(\cos \epsilon - \frac{\lambda}{L_b} \right). \quad (48)$$

Here the arc cosine is to be evaluated in the interval $(0, \pi)$. When $\cos \epsilon + (\lambda/L_b) > 1$, Δ_2 and Δ_3 have no real values, so that the δ -interval is a simple one, (Δ_1, Δ_4) .

The Δ 's are plotted in Fig. 10 as functions of ϵ , along with deflection curves for typical spacings s of the null line gratings. They correspond to the numerical case $L_b = 100\text{A}$ [L_b being approximately the thickness T of the crystal in Fig. 6(a)] and $\lambda = 0.05\text{A}$.

For normal incidence, the upper limit for δ is approximately $(2\lambda/L_b)^{\frac{1}{2}}$, from Eq. (48). Thus, in the numerical case, a diffraction from the given set of plane gratings is extinguished if it corresponds to a deflection greater than 0.03 rad, or to a null line spacing of less than about 1.7A. For non-normal incidences, however, Fig. 10 shows that the permissible deflections from the particular set of plane gratings is larger. It is more than doubled for an incidence 0.03 rad from normal. When we pass 0.032 rad for ϵ , the upper limit for δ keeps increasing, but the lower values begin to drop out.

Each deflection curve for the individual plane grating diffractions is intersected in two points (indicated by circles in Fig. 10) by Δ -curves. Solving Eq. (39) simultaneously with Eq. (46), (47) or (48), we find that these intersections have a common ordinate δ_1 given by

$$1 - \cos \delta_1 = \frac{1}{2} [(\lambda/L_b)^2 + (s/L_b)^2], \quad (49)$$

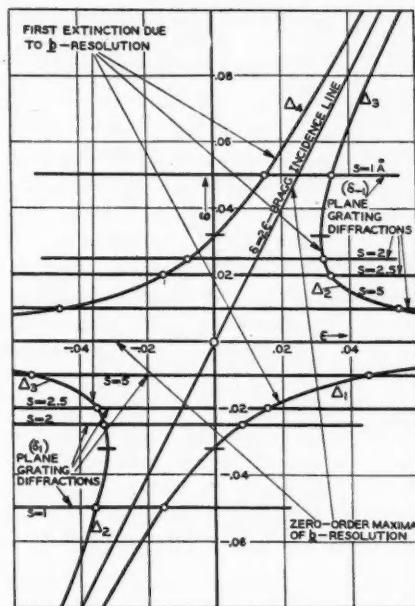


FIG. 10. Variations in incidence angle, $\frac{1}{2}\pi - \epsilon$, and deflection angle, δ , of space grating diffractions.

or, approximately,

$$\delta_1 = (\lambda/s)(1 + (s/L_b)^2)^{\frac{1}{2}}. \quad (50)$$

Thus, by varying the incidence direction, the deflection corresponding to a spacing s of the null line gratings is seen to vary upward from the value λ/s by approximately $50(s/L_b)^2$ percent.

From the same pairs of simultaneous equations, we find the abscissas ϵ_2 and ϵ_1 of the aforementioned intersections to be

$$\epsilon_2 = \frac{1}{2}\delta_1 - \sin^{-1} \frac{(\lambda/L_b)}{2 \sin \frac{1}{2}\delta_1}, \quad (51)$$

$$\epsilon_1 = \frac{1}{2}\delta_1 + \sin^{-1} \frac{(\lambda/L_b)}{2 \sin \frac{1}{2}\delta_1}; \quad (52)$$

or, approximately,

$$\epsilon_2 = \frac{1}{2}\delta_1 - (s/L_b), \quad (53)$$

$$\epsilon_1 = \frac{1}{2}\delta_1 + (s/L_b). \quad (54)$$

The approximations come from Eq. (41) by virtue of the slight difference between δ^0 and δ_1 . By the same token, these abscissas are approximately symmetric with respect to the Bragg value $\frac{1}{2}\delta^0$. Consequently, the given diffraction has

a tolerance in ϵ of s/L_b (rad) on either side of the Bragg value.

The foregoing treatment could be extended to the problem of diffraction by thick crystals or by polished surfaces. In these cases, most of the contributions are by rays entering and leaving a crystal by the same face. [See Fig. 6(b)]. A slight complication arises from the fact that the contributing plane gratings G_1G_1 , G_2G_2 , ... have a limited extent in the direction of depth from the boundary face AC . This is an outcome of the limited penetrating power of the electron beam, and it means that now two Laue conditions must be modified to the form (23). The picture is hardly complete, however, unless account is taken of the effect of "inner potential" and of multiple scattering, which are especially important in grazed regions. These factors are widely considered in the literature.²¹⁻²⁷ For the first, see references 21 to 25; for the second, 21, 26 and 27.

Conclusions

In view of the foregoing considerations, it appears that two salient features of the electron

²¹ E. Lamla, "Theory of electron diffraction with regard to more than two beams," *Ann. Physik* **32**, 178-89; 224-41 (1938).

²² S. Kalaschnikow, "Determination of inner potential of crystals from electron diffraction," *Physik. Zeits. Sowjetunion* **9**, 81-8 (1936).

²³ E. Rupp, "The significance of half-order interferences in electron diffraction," *Ann. Physik* **13**, 101-10 (1932).

²⁴ M. v. Laue, "The diffraction of an electron wave at a single layer of atoms," *Physical Rev.* **37**, 53 (1931).

²⁵ K. Lark-Horovitz, et al., "The origin of 'extra' rings in electron diffraction patterns," *Physical Rev.* **48**, 101 (1935)

²⁶ G. P. Thomson, "Electron diffraction by forbidden planes," *Nature* **126**, 55-6 (1930).

²⁷ Reference 10, pp. 391-393.

diffraction patterns produced by random assemblages of plane gratings and those produced by thin three-dimensional lattice arrays of atoms are only slightly different. These are the extent and the diffuseness of the patterns.

The diffuseness of the former type of pattern is admittedly the greater of the two, for the deflection has all values from the Bragg value δ^0 to $(2\delta^0)^{\frac{1}{2}}$. Still, as is evident from the initial flatness of the curve for δ versus ϵ (Fig. 9) we must expect most of the contributions to be close to δ^0 . As a result, there would be a maximum of intensity at a certain distance from the central spot on the photographic plate, accompanied by a continuous, but initially steep, shading off in the outward radial direction. In the presence of interfering factors, it might be difficult to distinguish between this type of sharpness and that exhibited by a pattern of the latter type.

We have also seen that extensive patterns are to be expected from three-dimensional, as well as from two-dimensional gratings. This is especially true if we assume, in the case of a thin sheet of mica, local variations in the orientation of the laminae.^{28, 29} Here again, it would appear difficult to choose between two- and three-dimensional gratings as the cause of any given pattern.

This approach to the problem of interpreting electron diffraction patterns is not essentially new.^{12, 17, 29} It is hoped, however, that the formulation and organization of procedure outlined here is sufficient to aid further quantitative studies.

²⁸ Reference 10, p. 390.

²⁹ W. L. Bragg, "Diffraction of x-rays by two-dimensional crystal lattices," *Nature* **124**, 125 (1929).

WENT to the Royal Institution last night in hopes of hearing Faraday lecture, but the lecture was given by Mr. Pereira on Crystals, a subject of which he appeared to be master, to judge by his facility and fluency; but the whole of it was unintelligible to me. . . . There is a cheerfulness, an activity, an appearance of satisfaction in the conversation and demeanor of scientific men that conveys a lively notion of the pleasure they derive from their pursuits.—GREVILLE MEMOIRS.

Apparent Depth and the Virtual Caustic

ALBERT V. BAEZ
Stanford University, Stanford, California

IN performing the well-known experiment of finding the index of refraction of a medium by measuring the apparent and actual depths of an object within it with a traveling microscope, the student is often not aware of the fact that the apparent depth as seen along the normal to the surface is the limit of a variable, which can be made to approach zero by viewing along the direction of grazing, instead of normal, emergence.

A simple experiment for finding the index of refraction can be devised in which the apparent depth is a variable. This experiment is quickly performed, and it is instructive because the critical angle is used. Although requiring only a block of glass, it is about as accurate as the usual method that requires, in addition, a traveling microscope. The order of difficulty does not exceed that of a first-year college physics course.

It is well known that an object at an actual depth d inside a medium of index of refraction μ seems to be at an apparent depth $a = d/\mu$ when viewed along the normal to the plane surface of the medium. The apparent depth, however, can be anything from a to zero, if we permit other angles of observation. In Fig. 1 the rays OA_1 and OA_2 , leaving O obliquely, become A_1B_1 and A_2B_2 in the air. The approximate location of the image I can be found by extending these rays backward as shown. The set of all possible virtual images in

a vertical plane forms the caustic curve shown by the dashed curve. It is obvious that the limiting apparent depth is zero as we approach grazing emergence, and that the maximum displacement b (Fig. 2) of the projection of the image on the surface is such that $b = d \tan c$, where c is the critical angle.

This leads to an interesting and rapid way of determining the index of refraction of a thick piece of glass in the shape of a rectangular parallelepiped. The glass block B shown in Fig. 3 has a straight scratch s on its upper face, parallel to one edge, and another scratch d perpendicular to and intersecting s , on a vertical face. The block is placed on a horizontal piece of rectangular coordinate graph paper with a line t drawn on one of its lines. If the image of t as viewed from directly above, with the eye at D , is lined up with s , obviously s will be directly over t ; but if we view obliquely, as at E , we see s lined up with some other line such as t' , parallel to t . By viewing along the direction of normal emergence, one should put s directly over t , and by viewing along the direction of grazing emergence, should determine the limiting position of t' and draw this line on the paper. The distance between the lines t' and t is b , which from a previous argument equals $d \tan c$. To find μ , however, it is only necessary to lay off d along t and t' and draw the hypotenuse x as shown in Fig. 4. A comparison of Figs. 3 and 4 will make it

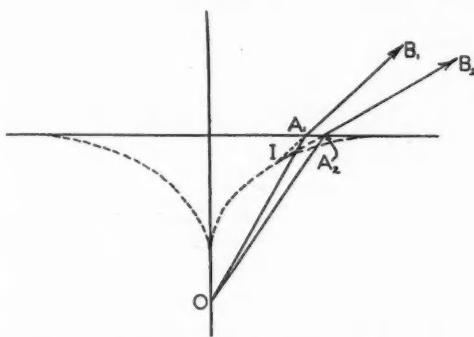


FIG. 1. The location of the image I when the object O is viewed obliquely. The dashed curve is the virtual caustic.

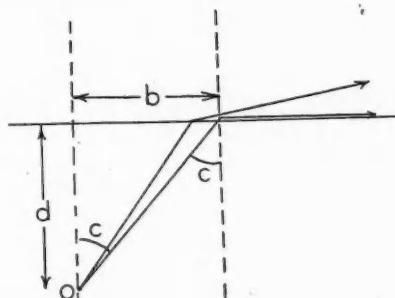


FIG. 2. As grazing emergence is approached, the apparent depth of the image approaches zero.

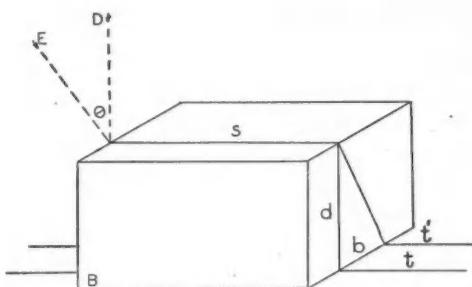


FIG. 3. In the special case of binocular vision in which a line joining the observer's eyes lies in a plane that is perpendicular to the plane refracting boundary, as θ approaches 90° , the image of that part of t' which lies under the glass approaches coincidence with s .

apparent that

$$x/b = 1/\sin c = \mu. \quad (1)$$

The value of μ can be calculated from the measured values of x and b . An error of 10° from grazing emergence will introduce an error of less than 3 percent in determining b and hence μ . This degree of accuracy will be found to compare favorably with that of the microscope method usually employed.

In binocular vision the images seem to lie on the virtual caustic only when the line joining the observer's eyes lies in a plane which is *perpendicular* to the horizontal refracting boundary, so that the foregoing discussion seems to be limited to this special case. However, the experiment can also be successfully performed when using binocular vision in the usual way, that is, so that the line joining the observer's eyes is *parallel* to the horizontal refracting boundary. In this case the image is not on the virtual caustic, but at a point directly above the object.¹ In either case, the limiting apparent depth is zero when viewing along the direction of grazing emergence.

Finally, if only one eye is used for viewing in a direction other than that of normal emergence, an astigmatic bundle of rays reaches it and the image might be at either of the two places mentioned, or somewhere in between. Specifically, the image could be somewhere on the line of the main emerging ray between its point of tangency to the virtual caustic and its point of intersection

with the vertical line through the object. The indicated experiment can be successfully performed with monocular vision or with either type of binocular vision described because *only the point of emergence of the limiting grazing ray is utilized*, and this is the same for all methods of viewing.

The Virtual Caustic

An alternative method to that usually given in textbooks on geometrical optics will now be used to deduce the equation of the caustic. The different positions of the image lie on the envelope of the one-parameter family of lines AB shown in Fig. 5. To use the familiar x, y notation, we orient the top face of the glass so that it is perpendicular to the plane of the paper and lies on the y axis. An object O is on the x axis at a distance d from the y axis. A ray such as OA bends to become AB in the air. The line AN is a normal to the face, and i and r have the usual meanings in refraction. We wish to find the equation of the envelope of the lines AB . This will be the locus of the images of O , and we will check by observing that when $i=0^\circ$, the image is at $(-d/\mu, 0)$, and when $i=90^\circ$, the image is at $(0, d[\mu^2-1]^{-1})$.

Consider the line AB . Its slope is $\tan i$, and its y intercept is $d \tan r$; hence its equation is

$$y = (\tan i)x + d \tan r. \quad (2)$$

In order to exhibit $\tan i$ as the explicit parameter, we effect a series of trigonometric transformations starting with $\sin i / \sin r = \mu$ and ending with $\tan r = \tan i [\mu^2 + \tan^2 i (\mu^2 - 1)]^{-1}$. Letting p stand for $\tan i$, we may write

$$y = px + pd[\mu^2 + p^2(\mu^2 - 1)]^{-1}. \quad (3)$$

Observe that as $p \rightarrow 0$, $y \rightarrow 0$. Dividing both members of Eq. (3) by p , we see that as $p \rightarrow \infty$, $x \rightarrow 0$;

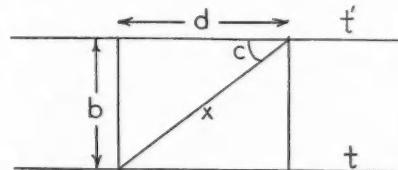


FIG. 4. By placing the block on its side, the actual depth d can be laid off along t and t' . The lengths b and x are the only two needed to determine the index of refraction.

¹E. R. Laird, *Am. J. Physics (Am. Phys. T.)* 6, 40 (1938); H. M. Reese, *ibid.* 6, 164 (1938); L. E. Kinsler, *ibid.* 13, 255 (1945).

hence the limiting rays are the x and y axes, respectively. Since $p = dy/dx$, Eq. (3) may be considered a differential equation of the Clairaut type, whose regular solution is the one-parameter family of lines we have been considering, and whose singular solution is the envelope we are seeking. The parametric equations of the envelope are²

$$\begin{aligned} y &= px + pd[\mu^2 + p^2(\mu^2 - 1)]^{-\frac{1}{2}}, \\ x &= \frac{\partial}{\partial p}(pd[\mu^2 + p^2(\mu^2 - 1)]^{-\frac{1}{2}}). \end{aligned} \quad (4)$$

These reduce to

$$\begin{aligned} y &= p^3 d (\mu^2 - 1)[\mu^2 + p^2(\mu^2 - 1)]^{-\frac{1}{2}}, \\ x &= -d \mu^2 [\mu^2 + p^2(\mu^2 - 1)]^{-\frac{1}{2}}. \end{aligned} \quad (5)$$

This parametric form is useful in plotting the graph of the virtual caustic. The choice of p as parameter has the advantage that physically p is the tangent of the angle of emergence. It is interesting to observe that while $p = \infty$ gives the exact position of the limiting image, $p = 10$ will

² Phillips, *Differential equations* (Wiley, ed. 3, 1934), p. 52.

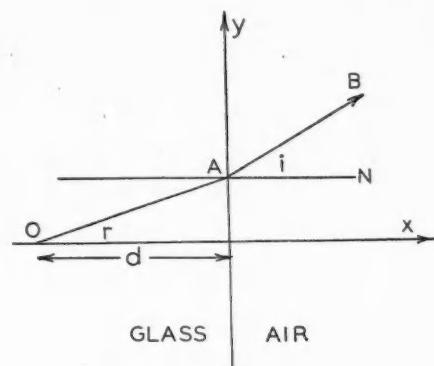


FIG. 5. The equation of the virtual caustic may be found by considering the envelope of the refracted rays AB .

give the location of the image with an error of less than 2 percent. Elimination of the parameter p yields the equation

$$(\mu x/d)^{\frac{1}{2}} + ((\mu^2 - 1)^{\frac{1}{2}} y/d)^{\frac{1}{2}} = 1, \quad (6)$$

which agrees with that given by Southall.³

³ Southall, *The principles and methods of geometrical optics* (Macmillan, 1910), p. 62.

Phase and Group Velocity

HERBERT JEHLE
Harvard University, Cambridge 38, Massachusetts

FIGURE 1 represents a boat, moving with velocity v_s , at two successive moments, together with waves as they appear at these instants. Suppose that a man in a canoe paddles along a line at right angles to the waves, in a direction opposite to the direction of advance of these waves. The man soon realizes the difference between phase velocity and group velocity of the waves through which he has to paddle. He expects the first wave to hit his canoe hard, but the first wave, big as it may look, fades away before the eyes of the paddler, and a later wave gets the satisfaction of splashing the canoe. After a few waves have passed by, the canoe is through what is called the "wave packet."

The phase velocity v_p and the group velocity v_g (see Fig. 1) of a wave packet are given by the

equations

$$v_p = 2\pi\nu/(2\pi/\lambda) = \omega/k, \text{ and } v_g = d\omega/dk,$$

where ν is the frequency; λ , the wavelength; $\omega [= 2\pi\nu]$, the angular frequency; and $k [= 2\pi/\lambda]$, the wave number. Let us calculate the relation between these two velocities for surface water waves by using simple dimensional analysis.¹

Consider first gravity waves, in which gravitational forces and inertial reactions act together to bring about the wave motion. If we suppose the water to be very deep, we may suspect that only the gravitational acceleration g , the wave number k and perhaps the density ρ will enter into the formula for phase (or for group) velocity.

¹ For a more detailed calculation, see H. Lamb, *Hydrodynamics* (ed. 5), pp. 358, 433.

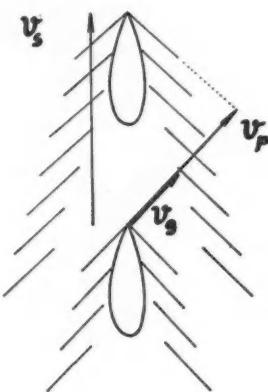


FIG. 1. Ship waves at two consecutive instants, a unit of time apart.

As the dimensions of these possible factors are $[g] = [LT^{-2}]$, $[k] = [L^{-1}]$, $[\rho] = [ML^{-3}]$, we can construct a velocity only by means of the combination

$$v_p = C(g/k)^{\frac{1}{2}},$$

where C is a dimensionless constant. Therefore,

$$\omega = kv_p = C(gk)^{\frac{1}{2}},$$

and the group velocity becomes

$$v_g = d\omega/dk = C \cdot \frac{1}{2}(g/k)^{\frac{1}{2}} = \frac{1}{2}v_p.$$

Figure 1 shows this relation. It is only in a rough approximation that ship waves can be considered as plane waves.²

We may study this phenomenon also by letting a stone fall onto a calm water surface. "If attention be fixed on a particular wave, it is seen to advance through the ring-shaped wave group, gradually dying out as it approaches the front, whilst its former place in the group is occupied in succession by other waves which have come forward from the rear."³

On the other hand, consider capillary waves, called "ripples," in which surface tension acts against inertial reaction forces. Here we may expect only the coefficient of surface tension T , the density ρ , and the wave number k to enter into the expression for velocity. As $[T] = [MLT^{-2} \cdot L^{-1}] = [MT^{-2}]$, $[\rho] = [ML^{-3}]$, $[k] = [L^{-1}]$, we can form a phase (or group) velocity

² Reference 1, pp. 409-413, gives a discussion of ship waves.

³ S. Russell, *British Association report* (1844), p. 369.

only by the combination

$$v_p = C'(kT/\rho)^{\frac{1}{2}},$$

where C' is another dimensionless constant. Therefore,

$$\omega = kv_p = C'(T/\rho)^{\frac{1}{2}}k^{\frac{1}{2}},$$

and the group velocity becomes

$$v_g = d\omega/dk = \frac{3}{2}C'(kT/\rho)^{\frac{1}{2}} = \frac{3}{2}v_p.$$

Here, the group velocity is larger than the phase velocity. These waves are observable as very small ripples on the surface of a calm lake, excited by a puff of wind.

Actually, in every wave both gravity and surface tension play a role. In that case,

$$v_p = (g/k + kT/\rho)^{\frac{1}{2}}.$$

We see that for long waves gravity is predominant, and we have $v_g \rightarrow \frac{1}{2}v_p$; whereas for short waves surface tension is the main force, and $v_g \rightarrow (3/2)v_p$. Let us study the intermediate region. We have

$$dv_p/dk = (-g/k^2 + T/\rho)/2v_p;$$

v_p therefore is a minimum for

$$\lambda = 2\pi/k = 2\pi(T/g\rho)^{\frac{1}{2}} = 2\pi(74/980 \cdot 1)^{\frac{1}{2}} = 1.73 \text{ cm};$$

the minimum value of v_p is $2^{\frac{1}{2}}(gT/\rho)^{\frac{1}{2}} = 23.2 \text{ cm/sec}$. Because of the formulas at the beginning of this article, $dv_p/dk = 0$ means $v_g = v_p$; v_g , however, has a minimum value of 17.8 cm/sec.

In both gravity and capillary water waves, the energy carried by the wave system moves with the group velocity. This gives the concept of group velocity a unique significance. The situation is, however, quite different in the case of electromagnetic waves in regions of anomalous dispersion.⁴

Some caterpillars demonstrate the propagation of a gravity-wave packet (Fig. 2). We have tried to teach one to move as surface-tension wave packets do, but so far without success.

FIG. 2. Caterpillar moving towards right.

⁴ See L. Brillouin and A. Sommerfeld, *Ann. Physik* 44, 177-240 (1914); *Congr. Inter. d'Electricité* 2, 1 (1932); Stratton, *Electromagnetic theory* (ed. 1), p. 334.

Algebraic Discussion of Lenses

ARON GURWITSCH

Harvard University, Cambridge 38, Massachusetts

IN most introductory textbooks, the proof of the minimum condition for the distance between object and image in the case of a converging lens is not given at all; and, where given, appears to be rather clumsy and awkward.¹ What follows is a simple proof of the fact that, if there is to be a real image in the case of a converging lens, the distance between object and image must not decrease below a certain minimum.² This proof offers a further advantage in that the equations permit the derivation of many properties of lenses, both converging and diverging.

It will be assumed that the lens is thin and that only the central portion is utilized. Denote by p the object-lens distance, by q the image-lens distance, by d the object-image distance, and by f the focal length. The sign convention adopted is as follows: p is measured from the object position to the lens, q is measured from the lens to the image position, d is measured from the object position to the image position. Each distance is positive if the direction in which it is measured is the direction in which the light is propagated. For a converging lens, f is positive; for a diverging lens, negative. The lens equation is

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{q}, \quad (1)$$

or

$$f = pq/(p+q). \quad (2)$$

Furthermore

$$d = p + q. \quad (3)$$

Elimination of p from Eqs. (2) and (3) yields the quadratic equation

$$q^2 - qd = -fd, \quad (4)$$

the solutions of which are

$$q = \frac{1}{2}d \pm (\frac{1}{4}d^2 - fd)^{\frac{1}{2}}, \quad (5)$$

or, briefly, $q = \frac{1}{2}d \pm D^{\frac{1}{2}}$.

¹ See, for example, the proof given by T. Preston, *Theory of light* (Macmillan, 1928), p. 114.

² A somewhat similar idea is expressed by J. P. Southall, in *Mirrors, prisms, and lenses* (Macmillan, 1918), pp. 239 ff., but his argument is not satisfactorily developed.

Instead of eliminating p , we may eliminate q from Eqs. (2) and (3) and obtain

$$p^2 - pd = -fd. \quad (4')$$

Evidently the right-hand member of Eq. (5) may be considered as an expression for either p or q in terms of d and f .

The argument developed thus far holds for all possible cases. We now proceed to the distinction between converging and diverging lenses.

Case I. $f > 0$, converging lens.—There are three possibilities.

(a) $p > 0, q > 0, d > 0$. The condition that Eq. (4) shall have real solutions is $D^{\frac{1}{2}} = \sqrt{\frac{1}{4}d^2 - fd} \geq 0$. Since $d \neq 0$, $d \geq 4f$. In other words, $d \geq 4f$ is the condition, both necessary and sufficient, for production of an image.

If $d = 4f$, Eq. (4) has only one solution and there is but one image; $p = q = 2f$.

For $d > 4f$, Eq. (4) has two solutions and there are two real images, provided $q > 0$ under all circumstances, that is, provided $\frac{1}{2}d > D^{\frac{1}{2}}$. To show that this is indeed true, let us assume the contrary. Then from $D^{\frac{1}{2}} > \frac{1}{2}d$ it follows that $-fd > 0$, which is obviously absurd. This proves the existence of conjugate foci.

This discussion permits us to interpret Eq. (4) in the following manner. Under the condition that $d > 4f$, the quadratic equation in question yields two solutions, Eq. (5), either of which may be considered as q , and the other as p . That the expressions for p and q are interchangeable is of course evident also from Eq. (1).

(b) $p > 0, q < 0, d < 0$. If $d < 0$ and $q < 0$, it follows from Eq. (3) that $|q| > |p|$. The converging lens functions here as a magnifying glass. However, in order that $q < 0$, it is necessary, in view of Eq. (1), that $p < f$. On the other hand, from $p < f$ it follows that $q < 0$. But, since $f > 0$, it follows from Eq. (2) that, when $q < 0$, $p+q = d < 0$ and hence, as already remarked, $|q| > |p|$. In other words, $p < f$ is the condition, both necessary and sufficient, that a converging lens shall function as a magnifying glass.

(c) $p < 0, q > 0, d < 0$. Since $p < 0$ and $d < 0$, but $q > 0$, it follows that $|p| > |q|$. In view of Eq. (1), $(1/f) - (1/q) < 0$; hence $q < f$. In this case, the real image is not only nearer to the lens than is the virtual object, but it is also located within the focal length. Conversely, if the real image is located within the focal length, it follows from Eq. (1) that $p < 0$. Consequently, in Case I(a) the position of the image must be outside of the focal length. In the case of a real image located within the focal length, in view of Eq. (2), $d < 0$ and also $|d/p| < 1$. Hence $|p| > |q|$.

The condition $\frac{1}{4}d^2 + f|d| > 0$ is *ipso facto* fulfilled in both Cases I(b) and I(c).

Since Eqs. (1) to (5) hold for all possible cases, we may again interpret Eq. (5) as yielding expressions for both p and q . However, since in Case I(b), $q < 0$, $p > 0$, and in Case I(c), $q > 0$, $p < 0$, we are not free, in contradistinction to Case I(a), to consider either solution as an expression for q or p . In Case I(b) the solutions of Eq. (4) are

$$\begin{aligned} p &= -\frac{1}{2}|d| + (\frac{1}{4}d^2 + f|d|)^{\frac{1}{2}}, \\ q &= -\frac{1}{2}|d| - (\frac{1}{4}d^2 + f|d|)^{\frac{1}{2}}. \end{aligned}$$

For Case I(c) we have

$$\begin{aligned} p &= -\frac{1}{2}|d| - (\frac{1}{4}d^2 + f|d|)^{\frac{1}{2}}, \\ q &= -\frac{1}{2}|d| + (\frac{1}{4}d^2 + f|d|)^{\frac{1}{2}}. \end{aligned}$$

Case II. $f < 0$, diverging lens.—There are again three possibilities.

(a) $p < 0$, $q < 0$, $d < 0$. In view of Eq. (1), which may be written as $-|1/f| + |1/p| < 0$, it follows that $|p| > |f|$. In this case, the condition $D \geq 0$ must again be fulfilled. However, the physical meaning of this condition is different from that in Case I(a), because there is here no screen to be placed at a proper distance from the real object. If a virtual object is given whose position is outside of the focal length, there is a virtual image so located that the distance between object position and image position fulfills the condition in question. In fact, $d = p^2/(p-f)$; hence

$$D = p^2(p-2f)^2/4(p-f)^2.$$

The same reasoning applies to Case I(a) if, instead of starting from d and seeking the conditions under which an

image may be obtained, one considers the actual real image that is produced at given positions of both object and lens.

The right-hand member of Eq. (5) may here be written

$$-\frac{1}{2}|d| \pm (\frac{1}{4}d^2 - f|d|)^{\frac{1}{2}}.$$

This member is evidently always negative. Its interpretation is the same as in Case I(a), except that both object and image are here virtual.

(b) $p < 0$, $q > 0$, $d > 0$. If the virtual object is located within the focal length,

$$q = \frac{pf}{-|p| + |f|} > 0.$$

Hence, in view of Eq. (2), $d > 0$, and, from Eq. (3), $|q| > |p|$. In this case Eq. (5) is

$$\begin{aligned} p &= -\frac{1}{2}|d| - (\frac{1}{4}d^2 + f|d|)^{\frac{1}{2}}, \\ q &= -\frac{1}{2}|d| + (\frac{1}{4}d^2 + f|d|)^{\frac{1}{2}}. \end{aligned}$$

The right-hand member of the second equation is evidently positive.

(c) $p > 0$, $q < 0$, $d > 0$. For a real object, Eq. (2) yields $q < 0$, but $d > 0$. It follows herefrom that $|p| > |q|$, in view of Eq. (3). Since, from Eqs. (2) and (3), $|f/q| = p/d$ and $d = p - |q|$, it follows that $|f| > |q|$. The virtual image corresponding to a real object is located within the focal length. The interpretation of Eq. (5) is here

$$\begin{aligned} p &= \frac{1}{2}d + (\frac{1}{4}d^2 + f|d|)^{\frac{1}{2}}, \\ q &= \frac{1}{2}d - (\frac{1}{4}d^2 + f|d|)^{\frac{1}{2}}, \end{aligned}$$

q obviously being negative.

In both Cases II(b) and II(c) the condition $D \geq 0$ is *ipso facto* fulfilled.

Some Advice to Students

THE question may arise in the mind of someone present—What may I do? Do not enter upon research unless you cannot help it. Ask yourself the "why" of every statement that is made, and think out your own answer. If through your thoughtful work you get a worthwhile idea, it will get you. The force of the conviction will compel you to forsake all and seek the relief of your mind in research work. You can prepare yourself for work. The paintings of the great masters, the compositions of great musicians, the sermons of great preachers, the policies of great statesmen, and the campaigns of great generals, do not spring full bloom from barren rock. Your training here is but a preliminary step in preparation for your life work. Mackenzie practiced 30 years before he

wrote his book on the heart. Training is required. As Osler says, "live in a day-tight compartment doing each day's work well." If you are a true student you will be more dissatisfied with yourself when you graduate than you are now. It is not within the power of the properly constructed human mind to be satisfied. Progress would cease if this were the case. The greatest joy in life is to accomplish. It is the getting, not the having. It is the giving, not the keeping.

I am a firm believer in the theory that you can do or be anything that you wish in this world, within reason, if you are prepared to make the sacrifices, think and work hard enough and long enough.—FREDERICK GRANT BANTING, in an address at Edinburgh University, 1928.

Librarian among the Physicists

ROBERT H. WHITFORD

Physics-Chemistry Librarian, College of the City of New York, New York 31, New York

NOT so many years ago, an unusual opportunity created drastic change in the familiar routines of a certain college science librarian. For convenience of reference he may be named Birtrow Hedforth, a cognomen quite obviously pseudonymous in character. To Birtrow was suddenly presented the chance to sever temporarily his 18 years' connection with the college library, in order to be associated with the department of physics in the same institution as a full-time teacher of introductory physics. As science librarian, he had long enjoyed the official status of instructional staff membership, in accordance with the enlightened and ever-growing recognition among educational authorities that librarianship is a teaching profession. However, disregarding a few lectures to students on library technics, he sorely lacked previous teaching experience in the narrower sense of actual classroom work.

Contemplation

The prospect of conducting class was therefore not without its terrifying aspects to one whose opportunities for facing an audience had been so limited. Moreover, retreat was unthinkable because the genuine need for additional physics teachers during the crucial period made acceptance of the new duties practically obligatory. Hedforth contemplated his new assignment with a buoyant enthusiasm somewhat tempered by latent misgivings. Would he prove to be a good teacher in every sense of the word, one able to lend helpful guidance and impart lasting knowledge without playing the pedant or disciplinarian? Would the possibility of incipient stage fright be insurmountable? How would he fare with the subject matter of physics itself with respect to propounding lucid expositions and furnishing valid replies to myriads of relevant questions? At any rate, he firmly resolved to exert every effort to justify the miraculous confidence that had been placed in a lowly librarian as a teacher of physics.

Concerning subject matter, incidentally, Birtrow was on reasonably safe ground. He had received a good technical education culminating in higher engineering and other degrees. But a long period indeed had elapsed since his fortuitous

acquisition of an unbroken string of A's in physics during college days, and he truthfully conceded that a current duplication of this feat would have demanded prodigious study. This need for review was felt keenly by others as well, resulting in the establishment of "refresher courses" at the college for those who were volunteering to teach a subject different from their accustomed specialties. The greater maturity and understanding of these individuals undoubtedly would contribute towards their deriving maximum benefit from participation in a review class. To the professor of physics who conducted such a refresher course attended by Birtrow will forever be directed the latter's sincere gratitude for the twofold bounty of inspiration and information received. Not only were sound principles and correct procedures patiently expounded, but a philosophy of physics teaching was created as well. Moreover, a rigorous universal method for expeditiously solving most of the basic problems of mechanics, whether in statics or dynamics, translation or rotation, was presented so lucidly that the subject matter of the course was unified. This integration served to render Birtrow's future teaching task much easier, while greatly facilitating the learning progress of his students. At the conclusion of the review course, he and the other prospective physics teachers received imposing diplomas in the shape of 4×6-in. form cards.

Action

So it was that Birtrow Hedforth sallied forth one bleak January day on his teaching mission. He had prepared his lecture topics and subject matter exceedingly well, wisely heeding his superior's friendly warning that students were quick to sense such shortcomings on the part of their teachers. He met his first two classes of the day's schedule and conducted them without disturbing incident, but uneasiness and lack of aplomb in the unfamiliar professorial role undoubtedly marred his presentations. Insidious doubt even entered his mind that he could carry on successfully, and that perhaps it might be prudent to withdraw from the whole undertaking before too late. But the third class marked the turning point, for thereafter misgivings gradually

disappeared and confidence steadily rose. Pre-occupation with unfolding the concepts of physics and with guiding the interplay of ideas in class discussions soon left no room for disturbing subjective feelings. Each additional day's smooth sailing heightened the new-found enjoyment of teaching, again proving the trite old adage about nothing succeeding like success.

Speaking of the passing time, perhaps Birtrow's most amazing discovery during his new work was the relative shortness of the conventional class period. One would suppose that defining an hour as one twenty-fourth part of the mean solar day should fix its magnitude with sufficient exactitude. This seems strongly belied by the apparent rapidity with which a teaching hour vanishes. The neophyte instructor usually prepares to deliver much more material than could possibly be covered, let alone intelligently absorbed. He soon finds that the minutes slip away fleetingly unless carefully apportioned among the various phases of classroom activity. Especially must he acquire the knack of guiding discussion into fruitful channels, and be able to sense the psychological moment for proceeding to a new topic without slighting the old. Under existing curriculums, the time that can be devoted to background literature is also not as boundless as the nonteaching librarian casually assumes, but this situation will steadily improve as more and more teachers become library conscious. Books like *Physics, the pioneer science*¹ and *Mechanics, molecular physics, heat, and sound*² represent noteworthy forward steps in this direction by the textbook writers. Critical re-examination of teaching objectives should accompany this trend towards greater rapprochement between classroom and library in order to eliminate superficial subject matter that fails to contribute towards valid educational goals. The time so gained may be utilized to cover fewer topics with greater thoroughness and unity, but especially to unfold the methodology of science and inculcate the scientific habit of thought. There should be ample opportunity to introduce cultural and historical material of the type regrettably omitted from Army physics classes because of pressure of time.

Comparison

What differences did Birtrow find between classroom teaching and library work? The teacher

experiences the exhilarating responsibility of being sole director of class activity. He plans, initiates and conducts the work of each hour in the manner best calculated to develop among his students sustained interest and lasting knowledge. These students he must learn to know by name, personality and achievement. He is their personal guide and mentor for an allotted period of 16 to 18 weeks of the college year, during which time they must be raised as a group from one level of intellectual attainment to another. Finally, the mutual success of teacher and students in accomplishing this goal is measured on the fateful day of reckoning at semester's close.

The librarian, on the other hand, does not have a finite group of students to work with, except in the broader sense of that portion of the student body frequenting the library. He tends to learn only the names of those who have focused themselves upon his attention as avid borrowers, persistent questioners, repeated delinquents, or the like. He usually does not initiate problems, but waits expectantly for them to be brought to him, whereupon he will do his utmost to unlock the literature for their solution. His helpfulness to a student questioner in any instance is very real and personal, but the chain of continuity is not as strongly welded as in the sustained teacher-pupil relationships of classroom teaching. The semester's end is less strenuous for the librarian than for the harassed teacher who is striving to evaluate his students' accomplishments in time to beat the office deadline for reporting of grades.

In the nature and length of working hours and schedules arises another major divergency. The librarian's work is spread rather uniformly throughout a full seven- or eight-hour period daily, whereas the teacher may undergo the physical strain of teaching three or four classes in succession, but then is unscheduled for the rest of the day. This freedom, however, is diminished by such exigencies as marking examinations, preparing lecture material and reviewing subject matter. These tasks become less onerous for the teacher who has taught the same course repeatedly, with the result that he has considerable spare time for general intellectual advancement through reading, research, writing articles, and so forth. He also enjoys longer summer vacations and more holidays. The librarian usually loses out on many such windfalls because the holiday proclamation ordinarily is "interpreted" as being relatively inapplicable to him. He cannot be

¹ L. W. Taylor (Houghton Mifflin, 1941).

² Millikan, Roller and Watson (Ginn, 1937).

accused of hankering for increased leisure, as the free time would be just as profitable for his intellectual advancement as for the teacher's. He deserves time off more comparable with that granted his classroom colleagues on the instructional staff. Moreover, if he were to find reasonable opportunity to peruse contents of books or journals in a manner contributory to his intellectual development, even during library hours, he should be able to do so with the free conscience of the teacher who habitually does the same on a more leisurely basis. This is in accordance with an enlightened administration's awareness that self-improvement of staff members enhances their utility to the institution served. Incidentally, the librarian might well be spared, possibly after five years' satisfactory service, the onus of filling out to the last quarter-hour an itemized time sheet for monthly meticulous scrutiny.

Random Reflection

Among Birtrow's many pleasant reminiscences of his teaching days are the kindnesses and support extended by his departmental chairman; the democratic friendliness and cooperation of his fellow teachers; the confidence shown by his supervisors in designating him to prepare a final departmental examination for hundreds of students; the pleasure derived from manipulating laboratory and lecture equipment at firsthand; the satisfaction of daily facing a critical audience and holding one's ground successfully; the added prestige accompanying a teaching role. Then again, there was the feeling of knowing a subject to the core—something imperfectly experienced

until actually teaching it—and the joy of declaiming it in stentorian classroom voice rather than subdued library whisper. Also freely enjoyed was the professorial prerogative of coyly interspersing jokes, traditionally laughed at heartily by all students who entertain the remotest idea of passing the course. But Birtrow's greatest personal satisfaction arose in favorable comments on his teaching by students who had long ago finished his course (thus excluding the possibility of "apple-polishing.")

Foremost among his relatively unpleasant memories of classroom teaching lies the task of assigning final grades. The fair-minded, conscientious teacher strives hard to apportion good grades equitably among those most deserving, but borderline cases inevitably plague him with difficult decisions. He also shares in the disappointment of the "F" student, humbly conceding a portion of the failure to be his own.

Second, the daily grind of teaching is a strenuous one, demanding a strong constitution and an even disposition. One is kept constantly on one's toes in anticipation of forthcoming assignments, as well as in evaluation of those which have gone before, all under the relentless pressure of an inexorable schedule.

* * *

Thus does Birtrow Hedforth contemplate in retrospect the contrasts between librarianship and classroom teaching. Oft 'tis said that the grass looks greener on the other side of the fence. Having browsed contentedly for a while in both fields, irrevocable choice between them would be extremely difficult.

The Truth Is Whole

Use is the aim of Science! give
again
A golden sentence to the faithful
pen—
Dwell not on parts! for parts con-
tract the mind;
And knowledge still is useless when
confined.
The yearning soul, enclosed in
narrow bound,
May be ingenious, but is ne'er pro-
found:

Spoiled of its strength, the fettered
thought grows tame:
And want of air extinguishes the
flame!
And as the sun, beheld in mid-day
blaze,
Seems turned to darkness, as we
strive to gaze;
So mental vigour, on one object
cast,
That object's self becomes ob-
scured at last.

ELIZABETH BARRETT BROWNING

Reproductions of Prints, Drawings and Paintings of Interest in the History of Physics

24. Caricature of Dalton Thanking Gerrit Moll for Defending English Science

E. C. WATSON

California Institute of Technology, Pasadena, California

THE interesting caricature reproduced in this article has been the subject of an extensive study by the late ERNST COHEN,¹ Professor of Physics at the University of Utrecht, Holland, the results of which were published in a paper entitled "Ein physikalisch-chemische Karikatur."² It has been discussed also by E. M. BROCKBANK in his recent publication of some of DALTON'S

unpublished letters.³ For more details regarding it than are given here, the reader is referred to these two publications.

As a result of his investigations, COHEN came to the conclusion that the drawing represents JOHN DALTON (1766-1844) thanking GERRIT MOLL (1785-1838), Professor of Mathematics and Physics at Utrecht, for his defense of English science. The story behind it, as reconstructed by COHEN, is essentially as follows. During the early years of the nineteenth century exact science was at a low ebb in England⁴ and attempts were being made to obtain government support for science, particularly in the form of pensions for worthy scientists. MOLL, who had gone to England to carry out investigations on magnetic induction and had met and formed a high opinion of YOUNG, JOHN HERSCHEL, KATER, BREWSTER, FARADAY, and others, anonymously published a pamphlet on *The Alleged Decline of Science in England by a Foreigner* (London, 1831) in which he argued the case for English science with both wit and wisdom. Shortly thereafter the government voted DALTON a pension of £150, and the award was announced at the annual meeting of the British Association for the Advancement of Science. The humor of assuming these events to be related apparently appealed to some one capable of effective use of the engraver's tools, and as a result it was preserved for us in this friendly caricature. Unfortunately, the name of the artist, who signs himself XYZ, has not been determined.



Caricature of Dalton thanking Gerrit Moll for defending English science.

¹ Announcement has just been made [*Science* 102, 348 (1945)] that Doctor Cohen was killed by gas poisoning in the notorious concentration camp at Auschwitz (Poland) early in 1944.

² E. Cohen, *Mitt. Geschichte Med. Naturw. Tech.* [4], 15 253-270 (1905).

³ E. M. Brockbank, *John Dalton, some unpublished letters of personal and scientific interest with additional information about his colour-vision & atomic theories* (Manchester, 1944).

⁴ For example, in 1830 Charles Babbage (1792-1871), Cambridge mathematician best known for his work on computing machines, published a book entitled *Reflections on the decline of science in England and some of its causes*. See in this connection also footnote 1 to the sixth article in this series of reproductions, *Am. J. Physics (Am. Physics T.)* 7, 187 (1939).

An Apparatus for the Measurement of Vapor Pressures

ALVIN W. HANSON
The Citadel, Charleston, South Carolina

THE apparatus described in *Mechanics, Molecular Physics, Heat, and Sound*, by Millikan, Roller and Watson,¹ for measuring vapor pressures by the static method is so difficult and tedious to construct and so subject to damage in handling as to discourage the use of this desirable method. To construct such a unit requires considerable skill and patience and many hours of labor. Ether and other such volatile and inflammable liquids are not likely to be investigated because the explosion hazard in sealing off a glass tube containing these liquids is too great. The apparatus described in this paper is designed to be more rugged, to be easily constructed, and to allow the use of most liquids without hazard.

Figure 1 is a diagram of the apparatus. A 200-ml, round-bottomed, short ring-necked boiling flask, containing about 100 ml of mercury, is provided with a three-hole rubber stopper to admit a thermometer (not shown in Fig. 1), an access tube, and a glass tube connecting the mercury pool in the flask to an open mercury manometer. The rubber stopper is firmly fixed in position with wire, and the stopper and tubes are sealed with red sealing wax. A continuous column of mercury extends from the open side of the manometer to the closed side and over a siphon bend into the mercury in the flask. The liquid whose vapor pressure is to be measured lies on the mercury pool.

In operation the flask is heated in a liquid bath. The liquid in the bath should have a higher boiling point than the liquid whose vapor pressure is to be measured. The bath temperature is raised until the liquid in the flask has boiled long enough to purge the air in the flask through the open access tube. The access tube is then closed by a screw compressor clamp. The level of the mercury in the open side of the manometer will be above the surface of the mercury in the flask, if the bath temperature is above the normal boiling point of the liquid in the flask.

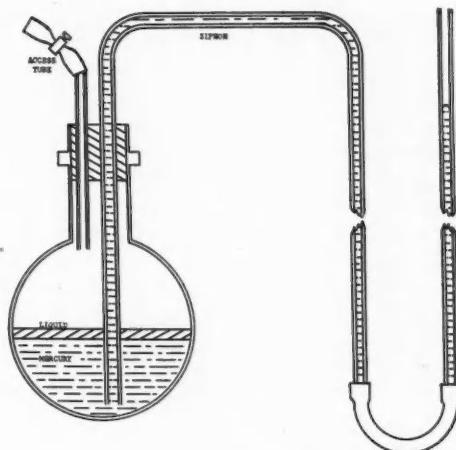


FIG. 1. Vapor pressure apparatus.

If the bath temperature is lowered below the normal boiling point of the liquid, the manometer level will be below the pool surface. When the temperature of the vapor has decreased to such a value that the vertical height from the top of the siphon to the open tube mercury level is just greater than the barometric height, the mercury column in the siphon will break. The pressure is then determined by the height of the mercury column in the tube rising from the mercury pool. This column is terminated at the top by a Torricellian vacuum, and atmospheric pressure no longer needs to be considered. The expression for the pressure before the siphon column breaks is $B - (R_L - R_t)$, and the pressure after the column breaks is $R_v - R_L$, where B is the atmospheric pressure, R_L is the average mercury pool level, R_t is the level in the open tube of the manometer and R_v is the level in the tube rising from the mercury pool. These levels are read on a meter stick mounted between the manometer tubes. The range of measurement is from 2 to 960 mmhg absolute.

The results obtained by a student are shown in Table I. These values are for temperatures

¹ (Ginn, 1937), p. 251.

TABLE I. Typical set of data.

Temperature (°C)	Values of vapor pressure of water (cmhg) Accepted	Experimental
0	0.46	0.45
10	.91	—
25	2.35	2.40
30	3.15	3.10
50	9.20	9.23
60	14.89	15.10
70	23.33	23.67
80	35.49	35.39
90	52.56	52.04

measured in the bath, as a thermometer was not placed in the vapor in this piece of apparatus.

Construction of apparatus.—A 114-cm length of 5-mm bore glass tube is first formed to make the siphon and the left-hand arm of the manometer. The first bend is 16 cm from one end. The second bend is in the same plane, 27 cm from the same end. Insert the access tube (an 8-cm length of glass tube), the thermometer and the siphon tube in the three-hole stopper. The stopper is to be high enough on the siphon tube so that the end of this tube will be about 0.5 cm from the bottom of the flask when the stopper is inserted. If accuracy is not desired, the temperature of the vapor may be determined from that of the bath, thus eliminating the need for a thermometer in the flask and for a three-hole stopper. To complete the manometer, join a 96-cm length of glass tube to the siphon tube by

an 18-cm length of rubber pressure tubing, wrapping the joints with wire. The support for the manometer is made from a board about 13 cm wide and 110 cm long.

The support for the flask is made from two pieces of wood about 20 cm long, of square cross section, the thickness and width being the same as the thickness of the board. One end of each piece is first mortised so that they may be rigidly fastened together in the form of an L by a screw. One end of this L-structure is then fastened across the top of the board; the other part projects downward, being parallel to and 7 cm away from the board, in position to hold the flask. A slot is cut in the flask support piece to receive the ring on the neck of the flask. Copper wire is wrapped around the neck of the flask and the wooden support when in assembled position. The stopper is secured in position with wire, and tubes and stopper are sealed with sealing wax.

To fill the apparatus with mercury, first pour mercury into the open arm of the manometer until the tubing is full to the upper bend in the left-hand arm. Clear this mercury of air by tapping the tube. Take a deep breath and pour again through the open arm at such a rate that the air in the siphon tube will be driven before the mercury into the flask. If at first you don't succeed, try again. Pour in enough mercury to half fill the flask.

IT is indeed wrong to think that the poetry of Nature's moods in all their infinite variety is lost on one who observes them scientifically, for the habit of observation refines our sense of beauty and adds a brighter hue to the richly coloured background against which each separate fact is outlined. The connection between events, the relation of cause and effect in different parts of a landscape, unite harmoniously what would otherwise be merely a series of detached scenes.—M. MINNAERT,
Light and colour in the open air.

Wire Models of Prisms

CLARK WERTENBAKER

College of the City of New York, New York 31, New York

WIRES models showing the shapes of various prisms and the paths of several rays through each have been found useful for demonstration and study. (See Figs. 1-6.) They are made with a standard aperture, and the ray systems are arranged so that when several prisms are put together to show the construction of a more complicated one, the rays match and can be traced through the whole combination.

The wire used is 0.080-in. galvanized iron; brass or copper wire conducts heat so fast that it is difficult to solder joints satisfactorily. Advance resistance wire is better than iron but is more expensive. The wire must first be straightened. This is done by attaching one end to a firm support and the other to the movable jaw of a heavy vise, and then closing the vise. The wire can then be cut into pieces of suitable length and bent to the proper shape in a vise, the corners being squared with a hammer.

The apertures are 7 in. wide and either square or rectangular. They should be made first and joined by the sides (see Fig. 7). The joint for an aperture should be made at the middle of one side. Handy and Harmon E-Z flow wire solder and flux are used. The ends of the wire are dipped

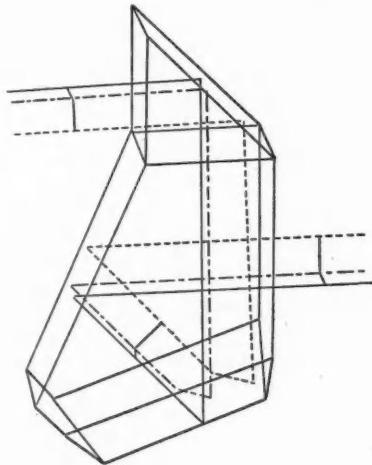


FIG. 1. Hensolt prism, an inverting system, consisting of a right-angle prism cemented to a penta prism. It is used in some field glasses. The penta angles are 112° , 112° and 90° .

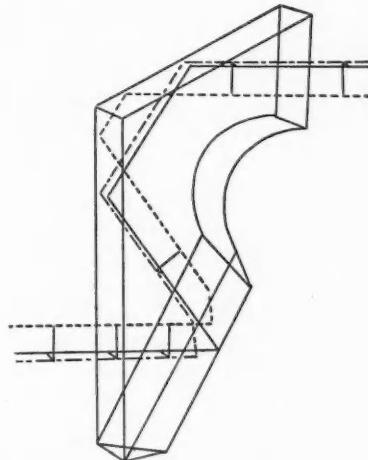


FIG. 2. Leman prism, an inverting system. The angles are 30° , 60° and 120° . The radius of the circular cut-out is equal to the aperture, that is, in the case of these models it is 7 in. This prism is used in machine-gun sights, and the axis of the gun may pass through the cut-out.

in the flux, which has been thinned with a little water, and wetted with the solder. A hand gas torch, in this case a National 3-A with N-1 tip, with a small hard flame furnishes the heat. Oxygen is not necessary. When the ends of the wire are butted together and heated, the joint sweats together without a fillet. It is possible to sweat four wires butted to a fifth if the fifth wire is heated very hot while it is held on the joint. Cardboard templets are convenient for bending the angles; 30° , 45° , 60° and 120° will be enough for most prisms.

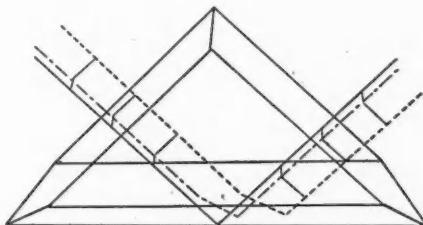


FIG. 3. Amici prism, an inverting system which deviates the rays through 90° . It is used in some periscopes and contour projectors.

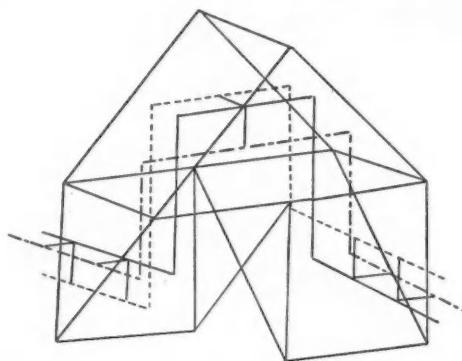


FIG. 4. Abbe prism, an inverting system, or Porro system of the second order. The angles are 45° and 90°.

Cross pieces in the apertures are used to support the wires representing the rays. The number and arrangement of rays is arbitrary. One may pass through the center of the aperture, with two others, above and to one side. They are bent at the proper angle when they reach the plane of a face of the prism. They can be supported by short transverse wires running from one to another or to the edge of the prism. In the present case these supporting wires were 2 in. long. In making crossovers for roof prisms, it is easiest to cut the wires at the faces of the prism and then connect the cut ends with a separate short piece. The

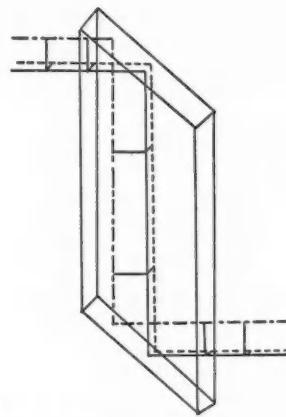


FIG. 6. Rhomboid prism, 45°. Its length may have any value, but generally is not more than twice the aperture or, in the case of these models, 14 in.

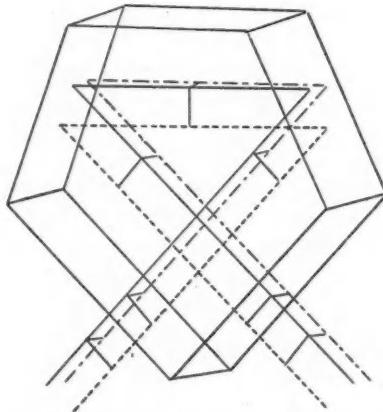


FIG. 5. Prandl, penta prism, or optical square which does not change the image but deviates the rays through 90°, whatever may be the orientation of the prism. It is used in range-finding systems. The angles are the same as for the Hensolt, penta prism, namely, 112°, 112°, 90° and 45°. The 45° angle is cut off, which makes the prism five-sided. Actually the angles are 112° 30'. In some instances, particularly in range finders, these angles may vary by a small amount for purposes of compensation.

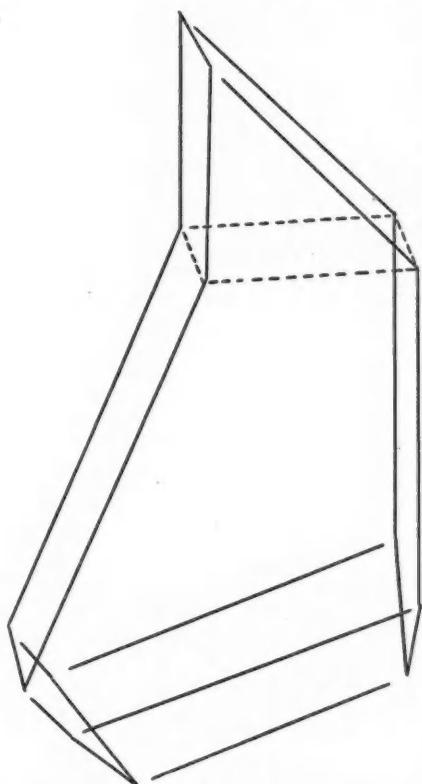


FIG. 7. Method of constructing the prisms (compare Fig. 1), showing which wires are bent and which are used as connectors. The dotted frame is one piece and is equal to the aperture.

angles formed should appear in projection as right angles when viewed at right angles to the aperture.

Painting the wires should be done after the model has been thoroughly washed in hot water to remove all traces of the soldering flux. The prism outlines can be painted white, the rays each a different color, with the transverse supporting wires black. It should be pointed out that in most cases the color has nothing to do with the color of the light that passes through the prism, and there is no dispersion in such prisms as are used in binoculars, for example.

It may be found more convenient to lay out the ray trace full size and then draw the prism around it, using this as a templet to form the side of the prism. This particular method was used in the construction of the Leman model (Fig. 2), as the wire had to be bent to form part of a circle.

The Photography of Wire Models

It is rather difficult to photograph wire models, especially when parts are colored, and still get sufficient contrast and perspective. Even with panchromatic plates the enlargements may have to be retouched. It was found that shadow projection worked very well, and it was possible to use anything from slow lantern-slide plates to news-print bromide for negatives.

Over a light-wooden frame, 45 in. square, was stretched a sheet of tracing cloth pulled tightly enough to be free of wrinkles. The model is suspended by means of a fine wire from a thumb tack on top of the frame.

At a distance of at least 30 ft (the limit in this case) and preferably more is placed a high intensity 85-w mercury arc with the aperture stopped down to about a $\frac{1}{4}$ -in. circle. This aperture should be in line with the center of the screen to within an inch or so. If stereoscopic pictures are to be made, the arc lamp should have a front guide made of a strip of wood so that the arc may be moved to the right or the left about a foot or so—the distance is not critical.

A carbon arc (pin arc) gives a much better image, but the illumination is so inconstant that exposures are difficult to handle. For a single picture it is better than the mercury arc, since it more nearly approaches a point source.

A camera is set up behind the screen and adjusted to the desired size of image, and the model is adjusted or guyed with additional wires until the proper perspective is achieved. The exposure time in the case of slow process plates at $f/8$ is about 45 sec. D-11 was used as the developer.

Stereoscopic photographs.—Without disturbing the camera, the light source is moved either to the right or to the left and another exposure made. At a distance of about 30 ft the source should be moved about 1 ft, though this distance is not critical. If more than one stereo is to be made, it is best to have this distance the same for all models. It is of some advantage to cast an additional shadow of a nail, piece of wax or some small object that does not interfere with the picture field to identify it as a stereo pair and which one of the pair.

Prints are made, and one is attached to a card of the proper size for the stereoscope with some cement that does not dry too quickly, such as rubber cement. The other print is swabbed with the cement and moved around in the stereoscope until the images fuse properly; then the whole is squeegeed down.

J. J. THOMSON was more than a great pioneer of science; he was a great teacher and an outstanding personality. If it was the fame of his scientific gifts which attracted to him at the Cavendish Laboratory students from every part of the civilized world, it was his personality which held them, and drew from them an affectionate reverence which was little less than worship. To them he was always "J.J.," and there was no one like him, and to them he devoted himself unspareingly. It was Thomson's deep conviction that the highest function of a university research department was not the multiplication of discoveries, but the training of men. In the universities and research departments of many lands, men trained by him in the arts of experimental research still carry on, in their various ways, the Cavendish tradition.—Journal of Scientific Instruments 17, 217 (1940).

NOTES AND DISCUSSION

An Undergraduate Student's Library

W. H. HARTWELL

University of New Hampshire, Durham, New Hampshire

AN undergraduate student interested in a professional career in physics should be encouraged to begin the collection of his own library. It will start with his textbooks, which he should be urged to mark and keep. Many students do not mark or make notes in their textbooks because they know that clean copies bring higher prices second hand.

The following list suggests some fundamental books, the first of which should be in the hands of every seriously minded student:

- L. W. Weld, *Glossary of physics* (McGraw-Hill, 1937).
- W. F. Magie, *Source book in physics* (McGraw-Hill, 1935).
- L. W. Taylor, *Physics, the pioneer science* (Houghton Mifflin, 1941).
- J. Strong, *Procedures in experimental physics* (Prentice-Hall, 1938).

The books suggested here are not beyond the reach even of students who are pressed for funds. The order of acquisition of others would depend on the student's interests—teaching, industrial research, and so forth. For example, any one interested in teaching would appreciate and want Sutton (Ed.), *Demonstration experiments in physics* (McGraw-Hill, 1938).

Sir Arthur Stanley Eddington—In Retrospect

HERBERT JEHLE

Harvard University, Cambridge 38, Massachusetts

ON November 21 it was a year since Sir Arthur Stanley Eddington died. He started his professional life as an astronomer at Greenwich Observatory, but soon afterwards moved to Cambridge where, as Plumian Professor of Astronomy and Director of the Observatory, he remained until his death. Some seven years ago he was elected President of the International Astronomical Union. It is a great hardship to the IAU that, after the war, it must resume its activities without Eddington, for he understood—not only as a scientist but as a colleague—how to bring together people separated by war. As International Secretary of the Royal Astronomical Society he undertook the courageous task of keeping up friendship between colleagues in universities and colleges everywhere, even through the dark years at the beginning of the war.

Eddington's scientific work was concerned with a wide range of problems. He started with an analysis of the dynamics of stars in globular clusters, and of stars and galactic clusters in our galaxy. From a study of the stability of galactic clusters moving in our galaxy, he concluded that the irregularly fluctuating forces acting on a star or cluster (exerted by neighboring stars or clusters) can, at present, be neglected in comparison with the smooth field of force due to the whole galaxy. This masterpiece of investigation brought stellar dynamics into its present state and formed the mathematical basis for the observational work of many explorers of our galaxy. He stimulated

and influenced research on the structure and the size of the universe.

The second field covered by Eddington was the investigation of the interior of the stars. Here the problem was to analyze the behavior of matter under extreme conditions of temperature and pressure, in particular, to give proper recognition to the role of radiation pressure. The theoretical explanation of the mass-luminosity relation was one of the fruits of this theory. Most stars were found to behave like spheres of ideal gas.

Another great piece of work was his investigations into the theory of relativity. He was the first to bring a knowledge of the theory to the English-speaking world. His *Mathematical Theory of Relativity* (1923) not only gave a survey of Einstein's special and general theories, but also introduced new ideas about the connection of the theory of gravitation with electromagnetic theory. Approaches on similar lines have been made in the great work of Weyl.

It is a fine thing—a tradition among British scientists—to give a version of what is going on in one's field, in a form readable by other people who are not familiar with the often quite complicated tools and language of the research man; moreover, in a form free from the disagreeable style which spoils so many a popularized presentation of science by suggesting that distorted motives of fame and power are at the root of research work. So Eddington wrote stimulating books such as *Stars and Atoms* (1927) and *Space, Time and Gravitation* (1920). But this searching mind could not be uninterested in the relations of its field of work with general knowledge. In his books, *The Nature of the Physical World* (1928), *Science and the Unseen World* (1929) and *The Philosophy of Physical Science* (1939), he helped to bridge the gap between these fields, thus stimulating the scientist's interest in philosophy and the philosopher's interest in science.

Eddington was not a collector of facts; he had a mind of vivid imagination. He knew what fields were most interesting, where the deepest problems lay. He worked hard, did the kind of work that others would not have done so soon, and took the risk of speculation.

He was keenly aware of the moral implications of scientific research. About his work we could say, with G. H. Hardy: "This subject has no practical use; that is to say, it cannot be used for promoting directly the destruction of human life or for accentuating the present inequalities in the distribution of wealth." Eddington stood for peace at a time when it was hard to take such a stand. He wrote in 1940:

I have found in the events of the last twelve months no ground for any weakening of my pacifist convictions. It has become clearer than ever that the choice is between pacifism and a drastic militarization of the nation (continuing in peace time) such as few would have contemplated two years ago. It is those who thought an intermediate course was practicable who have been disillusioned. They have seen that nowadays

war preparedness is a peril rather than a security, if it is not carried to such a pitch that the resources, the liberties, and the whole mind of the nation are surrendered to it as in the nations against whom we pit our strength. The faith of a pacifist must, I think, continue to rest on idealism or religious conviction rather than on any claim to political foresight; at least it is not for me (more familiar with the orderly way of stars than the vagaries of men) to say what would have been, and may yet be, the outcome of a pacifist policy.

Eddington, like Maxwell, Riemann and many other explorers, had an imaginative soul. Questioned about the existence of God, he wrote:

In the case of our human friends we take their existence for granted, not caring whether it is proven or not. Our relationship is such that we could read philosophical arguments designed to prove the nonexistence of each other, and perhaps even be convinced by them—and then laugh together over so odd a conclusion. I think that it is something of the same kind of security we should seek in our relationship with God. The most flawless proof of the existence of God is no substitute for it, and if we have that relationship the most convincing disproof is turned harmlessly aside. If I may say it with reverence, the soul and God laugh together over so odd a conclusion. For this reason I do not attach great importance to the academic type of argument between atheism and deism. . . . The crucial point for us is not a conviction of the existence of a supreme God but a conviction of the revelation of a supreme God, . . . the revelation implied in the indwelling of the divine spirit in the mind of man.

Terminal Velocity in Viscosity Measurements

M. ALDEN COUNTRYMAN

Illinois Institute of Technology, Chicago, Illinois

A RATHER common method of measuring the coefficient of viscosity η of a liquid in the sophomore physics laboratory is by means of the coaxial-cylinder type of viscometer designed by G. F. C. Searle¹ and manufactured by several laboratory supply houses. It is considered an "absolute" method in that the equation used,

$$\eta = \frac{(r_2^2 - r_1^2)gr^2m}{4\pi r_1^2 r_2^2 l v}, \quad (1)$$

may readily be derived from basic principles after making certain reasonable simplifying assumptions. These are: (i) the falling object of mass m which drives the rotating cylinder has reached terminal velocity v before measurements are made; (ii) no slippage takes place between the liquid layers adjacent to the rotating and stationary cylinders, respectively; (iii) the length l which appears in Eq. (1) is the "effective" length of the rotating cylinder, since the equation is derived without regard for end corrections; (iv) the motion of the liquid is nonturbulent.

In an effort to improve the apparatus and procedure for this experiment in our physics laboratory, it seemed

desirable to check these basic assumptions. The results of this experimental re-examination have been valuable to us, and they are presented here in the hope that they will prove useful to others.

Of the foregoing assumptions, the second and fourth seemed so reasonable that, from our standpoint at least, no experimental verification appeared to be justified. With regard to the third assumption, one possible method of eliminating the need for end correction, when oil is the liquid whose coefficient of viscosity is being measured, is to interpose a layer of water between the bottom of the stationary hollow cylinder and the oil; the very slight tendency of the oil to adhere to the water might then justify considering the lower layer—like the upper layer in contact with the air—as of negligible effect. This suggestion, though ingenious, has not as yet been tried out. Because of many secondary complications which suggest themselves, such as interfacial turbulence and similar pseudoviscous effects, it does not seem a very profitable solution. We feel that it is better for the student to make at least two sets of runs with two different lengths of rotating cylinder in contact with the oil² and thus eliminate the end correction by actual experiment.

To investigate the rate at which the driving masses approached terminal velocity and thus determine the initial distance which the student should allow a given mass to fall before timing its passage through a known distance, photographs were taken of several suspended masses, each falling under the conditions that actually exist during the experiment. Intermittent illumination by means of a Strobolux permitted photographing the falling object on 4×5-in. Portrait Panchromatic cut film; the photographs were subsequently enlarged to 11×14 in. for measurement. A meter stick was clamped approximately parallel to the path of fall and as near to it as possible so that the multiple images of the falling mass could be compared directly with the image of the meter stick and the distances traveled between successive flashes fairly accurately determined.

The results are given in Fig. 1, in which distance fallen is plotted as abscissa with corresponding average speed for each interval of time plotted as though it were the instan-

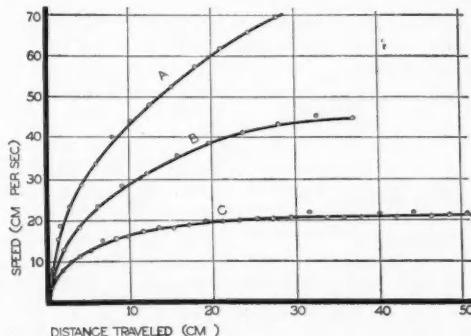


FIG. 1. Speed of fall in relation to distance traveled by suspended mass: A, 40-g mass, oil at 80°C; B, 40-g mass, oil at 75°C; C, 20-g mass, oil at 70°C.

taneous speed at the midpoint of the travel during that particular time interval. This is not of course strictly correct, but the error which results was not considered serious enough to justify attempts at correction. As the graph clearly shows, in the observed range of fall the only one of the three curves to indicate an approach to terminal velocity is *C* (20-g mass, oil at 70°C), and even in this case the speed appears to increase slightly after the suspended mass has traveled 45 cm from its point of release. Naturally, with the oil at higher temperatures and with a larger suspended mass, as in curves *B* (40 g, 75°C) and *A* (40 g, 80°C), failure to reach terminal velocity is much more evident. Unfortunately, the beam of light from the Strobolux was so small that it was not possible to observe how much farther the 40-g mass might have had to fall to reach terminal velocity, but even the meager data available clearly indicate that it is by no means safe to neglect the initial distance traveled while the suspended mass is being accelerated to terminal velocity. The procedure adopted in our laboratory is to use masses not larger than 25 g, and to start timing the descent as the falling mass passes a fiducial mark set 30 cm below the point of release.

¹ G. F. C. Searle, *Experimental physics* (Cambridge Univ. Press, 1934), pp. 250-254.

² The oil used in these experiments and in our laboratory was Texaco motor oil SAE 40, for which the refiners seem reluctant to give viscosity data.

Demonstration of the Peltier Effect

ROBERT W. KOZA

Georgia School of Technology, Atlanta, Georgia

THREE methods of demonstrating the Peltier effect which are described in current textbooks depend, respectively, on a transient Seebeck effect,¹ on a differential expansion of air masses surrounding a pair of junctions,^{2,3} and on a differential change in the resistance of coils wound around a pair of junctions.⁴

Among the desirable features of a demonstration apparatus of this type are a steady, quantitative indication of the effect, simplicity of design, and ease of construction with commonly available materials. It is also desirable that the indication be strikingly large without requiring the use of excessively large currents. This note describes a method of demonstration that seems to include these desirable features to a greater degree than do any of the aforementioned methods.

This method is an outgrowth of a demonstration seen at

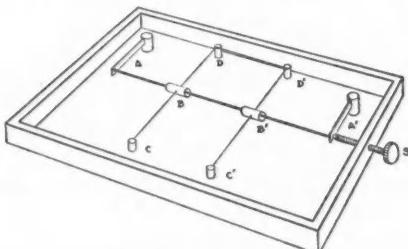


FIG. 1. Peltier demonstration apparatus.

the National Bureau of Standards several years ago. The principal apparatus used there was a piece of 60-40 gold-palladium wire welded between two pieces of 80-20 platinum-rhodium wire. When this composite wire was heated to incandescence by a direct current, one junction became visibly hotter than the average temperature of the wire, while the other junction became visibly cooler.

In the present method a similar composite wire is employed, but the difference in temperature between the two junctions is indicated by a thermocouple attached to radiation collectors which surround the junctions. The temperature of each radiation collector is proportional to the rate at which heat is produced in the incased junction. Heat is produced or absorbed at a junction at a rate which is proportional to the product of the Peltier emf and the direct current in the junction. If it is assumed that the average magnitude of the Peltier emf at the two junctions may be used for either one without appreciable error, the difference in temperature between the radiation collectors will be approximately proportional to the product of the average Peltier emf and the direct current in the composite wire. If the average temperature of the composite wire is constant, then the average Peltier emf will be constant, and the temperature difference between the radiation collectors will be approximately a linear function of the direct current in the wire.

The composite wire consists of a 5-cm length of No. 20 B & S gage constantan wire silver-soldered between two 5-cm lengths of copper wire of the same diameter. The ends of this composite wire are soldered to flat springs, *A* and *A'*, Fig. 1, and the wire is mounted under tension. Current is supplied to the composite wire through the springs and spring supports. The radiation collectors, *B* and *B'*, Fig. 1, are hollow cylinders, 3 mm in diameter and 10 mm long, made from 3-mil copper foil. These cylinders are suspended coaxially around the composite wire by the wires of a copper-constantan thermocouple, *CBDD'B'C'*, Fig. 1. Wires *CB* and *C'B'* are No. 30 B & S gage copper, and wires *BD*, *DD'* and *D'B'* are constantan wire of the same size. The junctions of this thermocouple are soldered to the radiation collectors. A galvanometer connected between *C* and *C'* indicates the temperature difference between the collectors. The composite wire, its supports, the radiation collectors, and the measuring couple are enclosed in a Lucite box to reduce the effects of air currents.

Any asymmetric Joule heating may be balanced out by a small axial displacement of the composite wire relative to the radiation collectors. The displacement is secured through the action of the screw *S*, Fig. 1.

The results obtained with this device are shown in Fig. 2. The dashed curve shows the deflection of the measuring galvanometer as a function of the direct current in the composite wire. The average temperature of the composite wire is determined primarily by the Joule heat produced, so that this temperature might be expected to increase considerably with increased current. Since the Peltier emf increases with an increase in temperature,⁵ the curvature shown is to be expected. If an alternating current is superimposed on the direct current in the composite wire and kept adjusted so that the sum of the two currents is

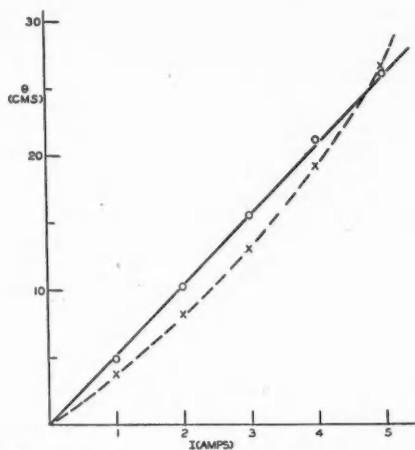


FIG. 2. Galvanometer deflection θ as a function of the current I through the junctions.

always constant, the average temperature of the wire and the Peltier emf will be approximately constant. The magnitude of the Peltier effect is then found to be approximately proportional to the direct current in the composite wire as shown by the solid line in Fig. 2. The deflections indicated in the figure were obtained with an ordinary L & N Type P wall galvanometer and a lecture demonstration scale placed about 1.5 m from the galvanometer.

¹ Wood and Cork, *Pyrometry* (McGraw-Hill, ed. 2), p. 34.
² Harnwell, *Principles of electricity and electromagnetism* (McGraw-Hill, ed. 1), pp. 181-182.

³ Sutton, Ed., *Demonstration experiments in physics* (McGraw-Hill, ed. 1), pp. 323-324.

⁴ Sudan, *Fundamentals of electricity and electromagnetism* (Van Nostrand, ed. 1), p. 271.

⁵ Reference 2, pp. 183-186.

Addition of a Third Term in the Lorentz Force Equation

AUSTIN J. O'LEARY
The City College, New York 10, New York

RATE of change of electromagnetic linear momentum has long been treated as an ordinary force in deriving the expression for electromagnetic momentum and in explaining radiation pressure. And in the restricted theory of relativity, force is defined as rate of change of linear momentum—mechanical plus electromagnetic. This broadened concept of force is quite familiar, yet the inclusion of an electromagnetic component as a legitimate part of the total force on a moving charge seems to have been generally overlooked.

In order to have a self-consistent theory of electrodynamics in which the same definition of force is used in all instances, it is necessary to add an electromagnetic force as a third term in the Lorentz force equation. In Gaussian units, the total force on a moving charge e is

$$\mathbf{F} = \mathbf{E}e + \frac{1}{c} \mathbf{ev} \times \mathbf{H} + \dot{\mathbf{G}},$$

where $\dot{\mathbf{G}}$ is the time derivative of \mathbf{G} , the electromagnetic linear momentum of the charge arising from its velocity \mathbf{v} in the presence of an external field; momentum due solely to the field of the charge itself is not included in \mathbf{G} because its rate of change would represent a self-imposed force, which has no meaning. The electric force and the magnetic force, comprising the first two terms, change the mechanical momentum of the charge, and the electromagnetic force changes its electromagnetic momentum. Though $\dot{\mathbf{G}}$ is small enough so that it may be neglected in most calculations, it must be taken into account in examples such as the one discussed in the next paragraph, where all terms involving first- and second-order powers of $1/c$ have been carried throughout the computation, along with those of zero order.

As an explicit example, let us apply the new expression for the Lorentz force to the interaction between two moving charges, making use of the results so carefully worked out by Page and Adams.¹ They have shown that in general, beyond a first degree of approximation, the mechanical momentum of the charges is not conserved by itself, and action is not equal to reaction; to be more specific, action is not equal to reaction when force is taken to mean rate of change of mechanical momentum only. From their Eq. (8) for the total electromagnetic linear momentum involved in the interaction, it may readily be seen that

$$\begin{aligned}\mathbf{G}_1 &= \frac{e_1 e_2}{2c^2} \left(\frac{\mathbf{v}_1}{r} + \frac{\mathbf{v}_1 \cdot \mathbf{r}_{12} \mathbf{r}_{12}}{r^3} \right), \\ \mathbf{G}_2 &= \frac{e_1 e_2}{2c^2} \left(\frac{\mathbf{v}_2}{r} + \frac{\mathbf{v}_2 \cdot \mathbf{r}_{12} \mathbf{r}_{12}}{r^3} \right).\end{aligned}$$

Letting \mathbf{F}_1 be the total force exerted on e_1 by e_2 , and \mathbf{F}_2 the total force exerted on e_2 by e_1 , we have

$$\begin{aligned}\mathbf{F}_1 &= \mathbf{E}_2 e_1 + \frac{1}{c} e_1 \mathbf{v}_1 \times \mathbf{H}_2 + \dot{\mathbf{G}}_1, \\ \mathbf{F}_2 &= \mathbf{E}_1 e_2 + \frac{1}{c} e_2 \mathbf{v}_2 \times \mathbf{H}_1 + \dot{\mathbf{G}}_2.\end{aligned}$$

There would seem to be no reason to expect that corresponding components of these two forces should be equal in magnitude or that they should be opposite in direction, and no such relation does exist except in special cases. It is a different story for the total forces. Page and Adams have demonstrated that the total linear momentum is conserved in accordance with expectation, from which it follows that

$$\mathbf{F}_1 = -\mathbf{F}_2.$$

And since the total angular momentum is also conserved, it is apparent that \mathbf{F}_1 and \mathbf{F}_2 are not merely parallel but are oppositely directed along the line joining the charges. Thus, according to the classical concept of force that is used in the consideration of radiation pressure, and according to the relativistic definition of force, we have this result: action and reaction are equal in magnitude and opposite in direction in all instances.

¹ L. Page and N. I. Adams, Jr., "Action and reaction between moving charges," *Am. J. Physics* 13, 141 (1945).

Demonstration of Optical Instruments

MARIO IONA, JR.

The University of Chicago, Chicago 37, Illinois

DISCUSSIONS of laboratory devices for an improved method of measuring the focal length of lenses when the image formed is a virtual image have recently appeared in this journal. The authors^{1,2} determine the position of the virtual image by using an auxiliary converging lens to form a real image of the virtual image on a screen and measuring the distance from this lens to a real object which is placed so that the auxiliary lens forms an image of it on the same screen.

This method of obtaining virtual image positions, which is equivalent to using an "eye,"¹ is recommended as providing another means of determining the focal length of lenses and of extending the range of object positions for which the focal length can easily be determined. Another important feature, however, is that it permits a satisfactory way of demonstrating the action of optical instruments by including the function of the eye.

The "final image" formed by optical instruments—for example, telescopes and microscopes—as ordinarily used is a virtual image. The eye is the optical system that forms a real image on a screen (the retina). In the demonstration of optical instruments one is therefore justified in attaching a system which is equivalent to the eye in order to obtain real images that can be observed by a large class.

Such an arrangement has been used for some time in our sophomore courses. Figure 1 shows an example of a

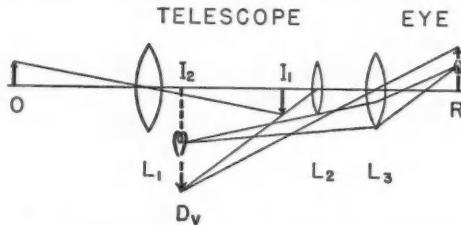


FIG. 1. Arrangement for demonstrating the images formed by the lenses of a telescope.

telescope arrangement. The objective lens L_1 forms a real image of the object O at I_1 which can be observed by interposing a ground-glass plate in the light path. The telescope eyepiece lens L_2 forms a virtual image I_2 , the position of which cannot easily be demonstrated unless the students look through L_2 individually. By using the additional system consisting of lens L_3 and screen R (retina), which together represent the eye, a real image is obtained on R when L_3 is properly focused. If a real object—for example, a carbon-filament lamp or other luminous object of distinct shape—is placed in the region D so that its light reaches R through L_3 (but not through L_2 !), a position D_v can be found at which both the virtual image I_2 of O and the auxiliary object form distinct images on R . Thus, the position of the final image (I_2) formed by the telescope is located, and the distance to which the eye must

accommodate is shown. The method allows one to obtain also the magnification by measuring the size of the image on the screen R with and without the telescope lenses. This is particularly convenient if the distance $I_2 L_3$ is sufficiently large that no change of focus of the "eye" is necessary. If the "eye" has to be refocused, the (angular) magnification can still be obtained from the size of the images on R and the distance $L_3 R$.

Similar arrangements can, of course, be used for the demonstration of the microscope and, with proper modifications, of the opera glass; they will also indicate whether the images are erect or inverted.

¹ P. K. Taylor, *Am. J. Physics*, 13, 167 (1945).

² J. P. Harper, *Am. J. Physics*, 13, 204 (1945).

A Classical Experiment Illustrating the Notion of "Jerk"

P. LE CORBEILLER

Crust Laboratory, Harvard University, Cambridge 38, Massachusetts

IN an article under this title¹ I have discussed the well-known "weight and threads" experiment, making the assumption that either thread would break under a tensile force equal to twice the weight of the suspended mass. It may be of interest to extend the discussion to all possible values of the ratio of the breaking force F to the weight mg .

Let $mg = \alpha F$, where α may have any value between 0 and 1. It is then found that the discussion depends upon the two equations

$$(\sin \theta)/\theta = \alpha,$$

$$[(1 - \cos \theta)/\theta]^2 + [1 - (\sin \theta/\theta)]^2 = \alpha^2,$$

which replace equations in reference 1 having the same left-hand member, and $\frac{1}{2}$ and $\frac{1}{3}$ on the right, respectively.

Figure 1 gives the result of the discussion. We have plotted vertically the ratio of S , the rate of change of the

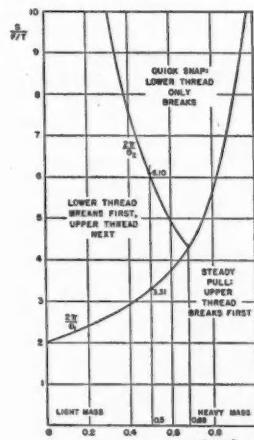


FIG. 1. Different types of outcome of the "weight and threads" experiment. Symbols: F , breaking force for either thread; T , period of oscillation of mass m suspended to upper thread; S , force applied to lower thread at time t .

applied force, to F/T , the ratio of the breaking force to the period of oscillation of the suspended mass. In reference 1, α was taken to be 0.5, and there were three possible cases. Figure 1 shows that for $\alpha < 0.684$ there are still three cases, but when $\alpha \geq 0.684$ there are only two. The corresponding critical value of the ordinate is 4.33.

¹ Am. J. Physics 13, 156 (1945).

The Moon Illusion

H. M. DADOURIAN
Trinity College, Hartford 6, Connecticut

In a paper¹ on the difference of the apparent size of the moon at the horizon and at the zenith, Professor Edwin C. Boring rejects the theory of the illusion given by Ptolemy and generally accepted since his time. After describing experiments with mirrors and artificial moons, he arrives at the following conclusion:²

There is no satisfactory further theory for explaining the phenomenon. It is not due to physical causes outside of the visual mechanism. It is not due to the greater brightness of the moon in elevation, when atmospheric haze is diminished. *It depends on raising or lowering the eyes.* Movements of the head, neck and body do not cause it. [The italics are mine.]

The arguments on which Professor Boring rejects Ptolemy's theory are very unconvincing to me; and his theory based on raising or lowering of the eyes is contrary to my experience and observation.

Ptolemy's theory may be put briefly as follows. The celestial dome appears as an oblate spheroid with a longer horizontal radius; the moon appears to be on the inner surface of the dome and, consequently, farther away from the observer when it is at the horizon than when at the zenith; but since the moon subtends very nearly equal solid angles in the eye of the observer at all elevations, and since of two objects having equal angular measures the one farther away appears larger, the moon appears larger at the horizon; furthermore, the celestial dome appears as an oblate spheroid because comparison with intervening terrestrial objects makes the radius of the horizon appear greater than the empty distance to the zenith. Boring's main reason for rejecting this theory is given in the following paragraph:³

Yet the argument is wrong for the simple reason that the moon in elevation looks farther away than the moon on the horizon—to all who have observed the moon much and thought about the matter. If asked why, they say, "The moon is so much smaller in elevation; of course, because it looks farther away"—thus inverting Ptolemy's logic.

A person to whom the moon appears to be farther away at the zenith than at the horizon should have one of the following experiences: (1) the celestial dome should appear to him as a prolate spheroid with a longer vertical radius, not as an oblate spheroid with a longer horizontal radius; (2) the zenith moon should appear to be beyond the apparent surface of the dome, in a hole in the sky; (3) the horizon moon should appear to be suspended somewhere between him and the apparent surface of the dome. I doubt if anyone has any of these experiences.

There is a sun illusion, similar to the moon illusion. Owing to the location of my house and to the fact that the setting sun can be observed more often than the rising moon, I have had many more opportunities to observe the horizon sun than the horizon moon. The house is located on Trinity College campus, which is on a north-and-south ridge. The sun sets over a parallel and higher ridge about 5 mi away, called Talcott Ridge. I have tried the following simple experiment to test Professor Boring's "raising or lowering the eyes" theory. I looked at the setting sun with naked eyes and, without raising or lowering my eyes, I interposed an overexposed film between my eyes and the sun. The result was, every time, a reduction of the apparent size of the sun to that of my impression of the zenith sun, accompanied by a corresponding reduction in its apparent distance. When I removed the film from the line of sight, the sun assumed its inflated size and greater distance. I tried the same experiment with the rising moon and obtained similar results.

As Professor Boring states, the horizon moon appears smaller when one bends down and looks at it between his legs or looks at it lying on his back with his head hanging down. This is the case also when the object viewed is a distant terrestrial object, a mountain range, for example. It is interesting to note that the color of a mountain range becomes lighter blue when viewed from a non-erect position.

A distant terrestrial object also appears smaller when one looks at it straight ahead while lying on one side. If one looks at a mountain across a valley, the mountain looks higher if the valley is visible than if it is not. Furthermore, the difference in apparent height is too great to be accounted for by the depth of the valley. When the lower part of a mountain is covered with a bank of cloud, it appears higher than when the cloud is not there.

There is another aspect of the illusion under consideration which is not discussed by Professor Boring. To me the zenith moon and the zenith sun appear to be about equal and constant in size, independent of the time and the place of observation. On the other hand, the horizon moon and the horizon sun appear to vary considerably in size. This variation takes place not when the eyes are raised or lowered, but when the image of the moon or the sun is accompanied by the images of different terrestrial objects in the foreground.

A number of times, I stood on the Trinity College Ridge and looked at the sun setting behind Talcott Ridge and observed the following phenomenon. When I focused my attention on the sun alone, it appeared about twice the size of my impression of the zenith sun. But when I focused my attention on the irregularities of the ridge, over which the sun was setting, as well as on the sun, the latter appeared much larger.

The foregoing considerations lead me to the following conclusions. (1) A distant object, whether celestial or terrestrial, appears larger when its image is accompanied by the images of intervening terrestrial objects. (2) This is due to conscious or subconscious comparison of the former with the latter. (3) If the change in the apparent size takes place on raising or lowering of the eyes, the primary cause is not

the change in the position of the eyes either in the skull or relative to the horizon, but to the fact that the objects in the foreground become more or less clearly visible and, consequently, comparison with them becomes more or less effective. (4) Ptolemy's theory needs a slight modification—the appearance of the celestial dome as an oblate spheroid is not the primary cause of the moon illusion. Both the illusion and the appearance of the celestial dome are due to the same cause, namely, the presence or absence of terrestrial objects in the line of sight.

¹ E. C. Boring, *Am. J. Physics* 11, 55 (1943).

² Reference 1, p. 59.

³ Reference 1, p. 56.

A Model to Demonstrate Spherical Aberration of a Convex Spherical Mirror

F. R. HIRSH, JR.

University of Southern California, Los Angeles 7, California

THE previously described model¹ for demonstrating spherical aberration in a concave spherical mirror may be modified to demonstrate the same phenomenon for a convex spherical mirror. Figures 1 and 2, together with reference 1, will aid in understanding the construction and operation of the model.

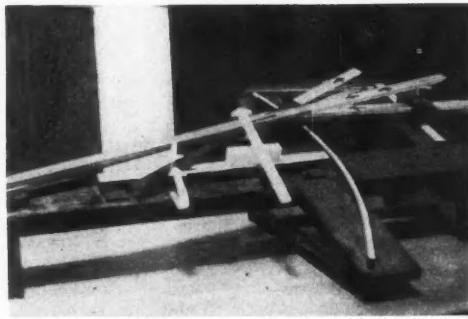


FIG. 1. Position of image and focal point for incident rays close to the axis.

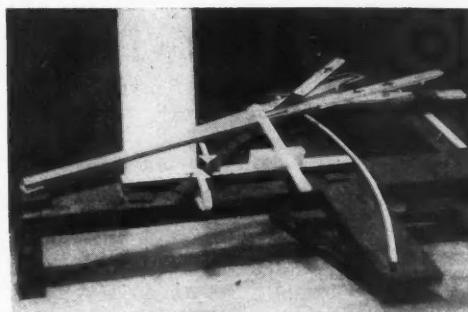


FIG. 2. Position of image and "focal point" when the parallel incident ray is far from the axis.

Two rays are represented in the model: one is parallel to the principal axis, and is reflected from the mirror as if it came from the "focal point;" the other is directed towards the center of curvature, and is reflected back on itself. The two rays may easily be distinguished by the black arrows on the white slats at the right-hand side of the model.

Again, as in reference 1, an isosceles triangle plays a part in explaining the shifting of the "focal point" as the incident parallel ray is moved out from the principal axis. The triangle lies at the left of (behind) the central section of the mirror; the base of the triangle is the radius of curvature drawn to the point of incidence of the parallel ray; the two equal legs are (i) the mirror principal axis between the center of curvature and the "focal point," and (ii) the reflected parallel ray (projected backwards) between the point of reflection and the "focal point." The operation of the model increases the length of the equal sides of the isosceles triangle, whereupon the "focal point" shifts away from the center of curvature and towards the mirror section.

The small white index arrow that points upward marks the point $(-\frac{1}{2}r, 0)$, which is the focal point for paraxial rays. The similar index arrow pointing downward represents the "focal point" for any particular position of the incident parallel ray. The large white arrow, pointing upward, represents the image position. The model is 3 ft long and 2 ft wide.

¹ F. R. Hirsh, Jr., *Am. J. Physics* 13, 267 (1945).

A N isolated fact can be observed by all eyes; by those of the ordinary person as well as of the wise. But it is the true physicist alone who may see the bond that unites several facts among which the relationship is important, though obscure.—HENRI POINCARÉ.

DIGEST OF PERIODICAL LITERATURE

Physical Problems in the Manufacture of Surgical Dressings

Many problems for the physicist arise in the manufacture of adhesive tape, surgical catgut, plaster of Paris bandages and dressings. Demands for wider applications of existing products, for new products having special properties, for more efficient manufacturing methods, for closer control of quality—all continually pose new problems.

Adhesive tape—Before the war the basic ingredient in pressure-sensitive tapes was natural rubber, which fitted admirably to manufacturing and to most surgical and industrial uses. Today, various combinations of synthetic rubberlike materials are used, and there have been many problems in connection with building into tape the physical properties it should have for various purposes.

The sticky substance of an adhesive tape has both elastic and viscous flow properties. It is a mixture of a rubberlike material, rosin, waxes and a filler (powder having fine particle size). The rubberlike material imparts elasticity to the adhesive mass, the rosin makes it soft and sticky, the waxes temper the stickiness, and the filler modifies and controls the elasto-viscous quality of the mass. In general, a good tape is one that can be pulled off cleanly from the surface to which it was applied; it should stick quickly, and not smear or creep.

The performance of the tape depends upon two basic physical properties: the modulus of elasticity E , and the relaxation time T . The modulus E is the force per unit stretch within the body. The relaxation time T is a measure of the rate of decay of the force required to maintain a fixed amount of stretch. This force decays in an exponential manner, and T is the time required for it to decrease to $1/e$ of its initial value. The coefficient of viscosity η is related to the elastic modulus and the relaxation time.

When the elastic modulus of the adhesive mass is too small, the tape cannot be pulled cleanly from the surface. When the relaxation time is too large, it is difficult to apply the adhesive to a surface and make it stick. When both E and T are small, the adhesive mass has too strong a tendency to creep. This tendency is often noticeable when tape is placed over a flexing muscle; the cloth backing tends to pull away from the sticky substance, leaving the latter to extend beyond the edges of the cloth. When E is too large, there is likely to be marked skin irritation because the tape holds the outer layer of skin relatively rigid and most of the yielding to the stresses of bodily movement takes place subcutaneously. In extreme cases, the outer layer of skin will come off. Thus, correct adjustment of the physical properties of the adhesive mass is very important.

In the study of elasto-viscous properties, use is made of an important equation,¹

$$\frac{dF}{dt} = E \frac{dS}{dt} - \frac{F}{t}, \quad (1)$$

where dF/dt is the time-rate of change of the stress in the material, and S is the relative deformation, or strain. In

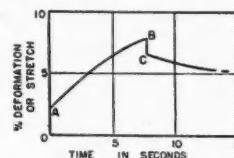


FIG. 1. Deformation of an elasto-viscous material.

the case of adhesives, conditions of use and manufacture are often such that dS/dt is constant; then Eq. (1) becomes, upon integration,

$$F = ET \frac{dS}{dt} + Ce^{-t/T}.$$

The term $ET \cdot dS/dt$ arises from the viscous forces within the material. The term $Ce^{-t/T}$ represents the elastic force induced by the deformation; its initial value is C .

Measurements of the physical properties of elasto-viscous materials are made with various types of plastometers and tensile strength testing machines. In one commonly used type of plastometer, a cylindrically shaped sample is compressed between two platens by a given force and the amount of compression in a selected time interval noted. In another type, the material is extruded through a small pipe by applying a very high pressure; this type resembles a capillary flow viscometer. Two other useful instruments are the penetrometer and the moving cylinder viscometer.² In the first, a needle penetrates the materials; in the second, the material is placed between two concentric cylinders, the inner cylinder having much the smaller diameter, and the motion of the inner cylinder for a given applied force is noted.

Figure 1 illustrates the characteristics of an elasto-viscous material.³ When a shearing force is applied, there is a sudden elastic deformation to A , and then stretching by a process of viscous flow until the force is removed at B . At this moment a sudden, small elastic contraction takes place to C , after which recovery continues over a long period of time. Usually the material never entirely recovers its original shape. The curve in Fig. 2 shows the relaxation of force in an adhesive mass. The adhesive was given a sudden fixed deformation, and the decay with time of the force needed to hold this stretch was noted. The relaxation time in this example was 0.45 sec.

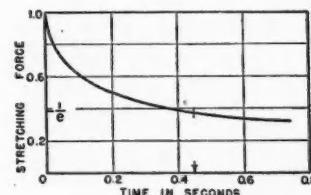


FIG. 2. Relaxation of force in a pressure-sensitive adhesive.

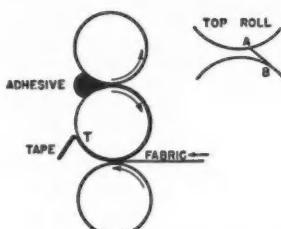


FIG. 3. Schematic diagram of a calender.

In handling the sticky materials used in making pressure-sensitive adhesives, the basic problem is to know how to put the material on its fabric backing and then how to let go of it. That is what a calender is designed to do. It consists of three heavy steel rolls mounted with their axes horizontal, one above the other (Fig. 3). The adhesive mass is placed between the top and center rolls, where it is pressed on to the center roll into a layer about 0.006 in. thick. This layer follows the center roll, and contacts the cloth at the nip between the bottom and center rolls. Enough pressure must be exerted at the nip to force the adhesive mass part way into the cloth; this is usually easy to do because the pressure can be adjusted over a wide range to fit the flow characteristics of the mass. The cloth and adhesive travel with the center roll for 80° and at T are pulled off as finished tape. The calender coats the tape in widths of 20 to 40 in. The desired narrow widths and lengths are cut afterward.

The calender must be able to release itself from the adhesive mass in two places: at the zone point where the top and center rolls begin to separate, and at the point T where the tape is pulled off the calender. The ability of the calender to let go at these two places depends upon roll temperatures, roll speeds and the properties of the adhesive.

The top roll is driven at a lower speed than the other two rolls so that the adhesive mass is stretched along the line AB (Fig. 3) as the center and top rolls separate. If the stretching along AB occurs with sufficient speed, the elastic force established will be large enough to pull the mass cleanly from the top roll. Equation (1) indicates how the calender should be operated to give the desired results with an adhesive material having a given set of properties. In a typical example, calculations based on Eq. (1) show that the combined viscous and elastic force established in the plane of AB is about 950 g per centimeter of width measured parallel to the axis of the roll. The elastic component of this force is about 200 g/cm, which was sufficient to separate the mass from the top roll.

When the finished adhesive is pulled from the calender, an elastic stretching of the adhesive again takes place in order to separate the mass completely from the calender. At the top roll nip, the rate of elastic stretching can be varied by changing the speed ratio of the top and center rolls, but at the pull-off point T the rate of stretching can be varied only by changing the angle at which the tape comes off the calender. Because of this, the range of properties of calenderable adhesives is restricted.

The amount of adhesive applied to the cloth must be closely controlled: a deviation of more than 5 percent from the correct amount results in tape of poor quality; also, the cost of the tape becomes excessive when too much adhesive is applied. A fluctuation of a few degrees in the temperature of the calender rolls or a change in the consistency of the adhesive will affect the thickness of the layer on the center roll. The calender operator must instantly be aware of such changes and must make adjustments in the calender.

To detect changes in its weight, the tape is passed between the plates of a condenser whose capacitance is being continuously measured and recorded. Any change in weight constitutes a change in the dielectric of the condenser and thus in the recorded capacitance. A glance at the capacitance chart will therefore warn the operator of any deviation in tape weight. This unusual application of condenser properties—its development was due to the Foxboro Instrument Company—has provided a much needed means of quality control and large reduction of waste in calendering operations.

Few realize how important adhesive tape is industrially, especially to the war effort. Most weapons and articles sent overseas must be wrapped to prevent corrosion by the action of rain, salt-water spray and sunlight; before being put to such use, the tape must pass rigid moisture penetration tests. The tremendous quantities of adhesive tape used for masking purposes in painting signs and insignia on airplanes and army vehicles must have special characteristics that adapt it to this use. The vast amounts of pressure-sensitive adhesive tape used in the electric industry must have high dielectric strength and must be free of any materials that would corrode copper wires when there is a voltage drop across the tape.

Surgical catgut.—Surgical catgut, like violin strings or tennis racket strings, is made from the lining of a sheep's intestine. It is used for internal stitching because the body is able to absorb the catgut after the wound has healed. Hospitals buy it packed sterile in sealed glass tubes, each containing a 60-in. strand. The strings are supplied in diameters ranging between 0.010 and 0.030 in.

Formerly, catgut was sterilized by packing thousands of tubes at a time, open end up, in an oil bath and heating to 155°C for at least 1 hr. The oil is a good heat distributor because its density decreases rapidly with rise in temperature, thereby setting up strong convection currents that carry the heat quickly and uniformly to all parts of the load of sutures. After sterilization, the oil was emptied from the tubes, and the tubes were partially filled with an antiseptic alcoholic solution, then sealed.

It was found that small quantities of the oil stayed in the catgut and, when the latter was imbedded in the body tissues, caused some local irritation and delay in wound healing. Since the oil could not be satisfactorily washed out, it was necessary to devise another sterilization process. This difficult problem of obtaining uniform and rapid heat distribution through a large mass of poorly conducting glass tubes was finally solved in such a way that all tubes of the load reach the same temperature simultaneously and there are no temperature gradients within the tubes.

The new sterilization method also presented a problem of

temperature control, and this was solved by using a set of thermocouples connected in series opposition. The arrangement used was able to anticipate a deviation from the desired sterilization temperature before any sensible change had taken place and to adjust the heat input so as to maintain a steady temperature. These two applications of thermal principles represent an important contribution to improved surgery.

Plaster of Paris bandages.—Large quantities of plaster of Paris bandages are used in making casts for the treatment of fractures and serious wounds. The usual bandage of this kind is made by impregnating crinoline with the dry plaster of Paris powder. As fast as the crinoline is impregnated with the powder, it is rolled up into 3- or 5-yd lengths. This type of bandage has many disadvantages. It is messy because the plaster of Paris falls out of it easily. When the bandage is wet out by immersing it in water, the plaster of Paris washes out of several of the outside layers; and the water penetrates very slowly to the center of the bandage, complete penetration requiring about 60 sec. Such a bandage makes a relatively poor cast.

A new bandage with none of these drawbacks has been developed. The plaster of Paris is fastened to the crinoline with a water soluble adhesive. The resulting layer of plaster has a rough surface; so, when the bandage is wound, there are small spaces between layers for water to enter. The bandage wets uniformly in 3 to 5 sec. It does not lose plaster either in the water or when handled prior to wetting. It produces a much stronger cast.

The manufacturing of these dustless bandages offered some interesting problems. The plaster of Paris is made into a thin paste which contains the ingredients for binding the dry powder to the crinoline. This paste must be applied to the crinoline in a comparatively short time, before the plaster has had time to harden, and the machine that applies it must be clean; if the plaster of Paris accumulated, it would harden and clog the spreading machine. Another difficulty is that this plaster of Paris paste does not flow freely, but must be pushed into place. It is a soft plastic solid which may be characterized by the Bingham equation for soft solids,⁴

$$\frac{dv}{dr} = \frac{1}{\eta}(F-f), \quad (2)$$

where dv/dr is the velocity gradient between center and walls of a capillary tube of radius r ; η , a coefficient of viscosity;⁵ F , the applied shearing stress; and f , the viscous shearing stress. Such soft solids will not flow unless F exceeds f .

The spreading machine developed for making this bandage depends for its operation upon the principles expressed in Eq. (2). It consists of a series of rapidly turning rollers to maneuver the paste into position on the crinoline, but the plaster of Paris cannot stick to the rollers because the

force pushing it on to the rollers is not large enough to overcome the viscous force. The spreading machine keeps itself clean.—HERBERT M. STRONG, Bauer and Black Co., Ohio State Univ. *Eng. Exp. Stat. News* 18, 28-33 (1945).

¹ J. C. Maxwell, *Phil. Mag.* [4] 135, 134 (1868).

² W. W. Pendleton, *J. Appl. Physics* 14, 170 (1943).

³ K. Bennewitz and H. Roetger, *Physik Zeits.* 40, 416 (1939); W. G. and J. M. Burgers, *First report on viscosity and plasticity* (Nordeman ed. 2, 1938); R. Houwink, *Elasticity, plasticity and structure of matter* (Cambridge Univ. Press, 1937); E. Jenckel and W. Holzmüller, *Zeits. f. physik. Chemie* A186, 359 (1940); H. Leaderman, *Textile Research* 11, 171 (1941); N. W. Taylor, *Rheo. Bull.* 12, 27 (1941).

⁴ E. C. Bingham, *Fluidity and plasticity* (McGraw-Hill, 1922). Chap. VIII.

⁵ The coefficient η differs from the ordinary coefficient of viscosity in that it must be measured by determining the rate of flow for at least two different rates of shear.

Check List of Periodical Literature

Vibrations in telegraph wires. F. Bauchwitz and R. N. H. Whitehouse, *Nature* 155, 243 (1945). Telegraph wires, coated with a layer of ice which was just beginning to thaw in the morning sun, were observed to be vibrating with a frequency of about 10 c/sec; there was no apparent wind or other evident cause. In an appended note, R. L. G. Gilbert, V. Robinson and W. N. Wilson suggest that the ice layer was noncylindrical; hence a very slight wind exerted a downward force on the wire if the latter were already moving downward. The motion was stopped by the elastic forces in the wire, whereupon the aerodynamic force also stopped. The wire then began to rise, in which motion the wind again helped. See Den Hartog, *Mechanical vibrations*, p. 343.

Science in liberal education. Brother I. Leo, *J. Chem. Ed.* 22, 187-193 (1945). In the Middle Ages science, represented by astronomy, was one of the seven liberal arts. Today there is a growing feeling that, though liberal education is a necessity, science should be relegated to a secondary place. A subject to be liberal must develop the mind, have disciplinary values, possess values of appreciation and must consider the ever-present problems of man. Judged by these criterions, science has a rightful place in a liberal education. Teachers of science are the ones who can do most to lessen the intensity of the controversy. In teaching, they should stress principles rather than facts; content rather than method; observation and reasoning rather than technics.

Density plummet thermometers. C. E. Lloyd, *Sch. Sci. and Math.* 45, 470-471 (1945). A set of three density plummets [*Am. J. Physics* 13, 59 (1945)], adjusted to sink in water at three different temperatures, say 18°, 21° and 24°C, constitute a rough but easily observed thermometer for a classroom. Alcohol can be used instead of water to reduce the time lag in response due to the high specific heat of water.

It requires a very unusual mind to undertake the analysis of the obvious.—A. N. WHITEHEAD.

RECENT MEETINGS

Kentucky Section

THE Kentucky Section of the American Association of Physics Teachers met at the University of Kentucky, Lexington, on October 27, 1945. Professor W. C. Wineland presided. The following papers were heard:

Thevenin's theorem. A. D. HUMMELL, *Eastern State Teachers College.*

Methods of measuring cannon velocities. HAROLD FLANNERY, *University of Kentucky.*

There'll be some changes made. G. V. PAGE, *Western State Teachers College.*

Acceleration-velocital magnetic forces. F. W. WARBURTON, *University of Kentucky.*

Ssh! scientists at work! D. M. BENNETT, *University of Louisville.*

The program was concluded with a brief business meeting and a luncheon at the Student Union.

Southern California Section

THE Southern California Section of the American Association of Physics Teachers met at Occidental College, Los Angeles, on October 6, 1945. Forty members and guests attended the morning session at which the following contributed papers were presented:

Several lecture demonstrations. L. E. DODD, *University of California at Los Angeles.*

Model to demonstrate aberration in a convex spherical mirror. F. R. HIRSH, JR., *University of Southern California.*

Dependence of the tuning of wind instruments on temperature. R. W. YOUNG, *USN Radio and Sound Laboratory, San Diego.*

Physics and a philosophy of teaching. G. FORSTER, *Pasadena Junior College.*

Surplus war equipment in the local area. (a) W. GEER, *University of Southern California;* (b) D. L. SOLTAN, *University of Redlands.*

The afternoon program consisted of an address on "Atomic fission," by Dr. William A. Fowler of the *California Institute of Technology*. This lecture, the first of a series of public lectures sponsored by the Section, was attended by more than 300 persons.

At the business meeting, R. W. McHenry reported the results of the first annual competitive physics examination for high school students, held at the University of California at Los Angeles on June 2, 1945. The examination was sponsored by the Section, and over 90 high school students participated. Certificates of award were given to the highest ten.

VERNON L. BOLLMAN, *Secretary-Treasurer*

New England Section

THE twenty-fifth meeting of the New England Section of the American Physical Society was held at the Springfield Division of Northeastern University, Springfield, Massachusetts, on October 13, 1945. At least 60 members were in attendance. The program included five short papers and three invited papers. There follow the titles of the invited papers and of those short papers that deal mainly with problems of teaching.

Experiments with uhf wave guides. G. F. HULL, JR., *Dartmouth College.* For the complete paper see *Am. J. Physics* 13, 384 (1945).

Jet propulsion and gas turbines for aircraft. A. F. YACOVONE, *Pratt & Whitney Aircraft Division, United Aircraft Corporation.*

Radar—principles and applications that can now be told. W. B. NOTTINGHAM, *Massachusetts Institute of Technology.*

Laboratory experiment on the measurement of time delay by use of

a cathode-ray oscilloscope. SHERWOOD GITHENS, JR., *Harvard University.* The complete paper will appear in a later issue of this journal.

Large currents through the glass envelope of an incandescent lamp due to electron emission from the filament. G. F. HULL, SR., *Dartmouth College.*

Phase and group velocity. HERBERT JEHLE, *Harvard University.* The complete paper appears elsewhere in this issue.

At the business meeting the following officers were elected for 1946: S. R. Williams, *Chairman;* H. Margenau, *Vice chairman;* G. F. Hull, Jr., *Secretary-treasurer;* N. C. Little, V. E. Eaton and R. D. Evans, *Program committee.* MILDRED ALLEN, *Secretary-Treasurer*

Western Pennsylvania Section

SOME 40 members and guests attended the meeting of the Western Pennsylvania Section of the American Association of Physics Teachers, held at Duquesne University on November 10, 1945. The meeting was opened with an address of welcome by the Reverend F. P. Smith, Dean of the College of Arts and Science, Duquesne University. The following papers were heard:

Speed and accuracy in computations, using approximation methods. A. G. WORTHING, *University of Pittsburgh.*

Resistivity apparatus for rod specimens. A. A. HAMMOND and C. WILLIAMSON, *Carnegie Institute of Technology.*

Social and political implications of atomic power. R. J. MAURER, *Carnegie Institute of Technology.*

Some crystallographic facts commonly misrepresented in physics textbooks. S. S. SIDHU, *University of Pittsburgh.*

The meeting closed with a luncheon and business meeting, and an inspection of demonstration equipment and the laboratories of the department of physics.

W. H. MICHENNER, *Secretary*

Oregon Section

THE Oregon Section of the American Association of Physics Teachers met at Willamette University, Salem, Oregon, on November 10, 1945. The following program was presented:

Loran: long-range navigation. K. TROLAN, *Oregon State College.*

Poggendorff's apparatus. W. WENIGER, *Oregon State College.*

Centripetal force apparatus. E. H. COLLINS, *University of Oregon.*

Impedance bridge. G. VASSALLO, *University of Portland.*

Educational demonstration program. CAPT. W. B. FORD, CAPT. R. E. REEVES, CAPT. M. R. HARRIS and SGT. J. C. JOYCE, *Army Air Corps, Area IX.*

Proposed reorganization of undergraduate studies. A. E. CASWELL, *University of Oregon.*

The physics involved in metallurgical processes. B. A. ROGERS, *N. W. Electro-Development Laboratory, Albany, Oregon.*

After the afternoon session a trip was taken through the alumina production plant of the Columbia Metals Company. The plant is being built by the Chemical Construction Company, by whose courtesy the trip was taken.

An extended discussion of the political and social aspects of atomic power led to the adoption of a resolution whereby the section expressed its approval of a statement made by some of the scientists concerned with the New Mexico project, to the effect that it is unrealistic to expect any nation possessed of the atomic bomb to agree not to use it in war, and that only a world authority, having sole control of the bomb, can enforce international law and insure peace.

E. HOBART COLLINS, *Secretary*